REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-

0659

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
November 22, 2000

3. REPORT TYPE AND DATES COVERED
Final Report, 04/01/97 - 12/31/99

4. TITLE AND SUBTITLE
A System Theoretic Approach to Autonomous Vehicle Dynamic Formation

5. FUNDING NUMBERS
G: F49620-97-1-0272

6 AUTHORS
Jason L. Speyer

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Mechanical and Aerospace Engineering Department
48-121 Engineering IV, Box 951597
University of California, Los Angeles
Los Angeles, CA 90095-1597

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NA
801 N. Randolph Street, Room 732
Arlington, VA 22203-1977

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Air Force position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

UL

13. ABSTRACT *(Maximum 200 words)*

A dynamic network of cooperating agents is characterized by a spatially distributed set of dynamic nodes or agents which coordinate to perform the mission objectives. These coordinated clusters could be composed of an array of satellites constructing a large aperture radar or a swarm of UAV's used to suppress enemy air defenses. These mission objectives are to be achieved in the presence of large uncertainties due largely to a hostile environment. Within this context, nodes may fail at various levels, measurements may be highly corrupted and communication channels may be severely limited due to jamming. Communication links are further challenged due to power constraints and large spatial dispersion, producing tradeoffs between uncertain information, latency and bandwidth constraints. A decision and allocation process appears computationally intractable, especially if mechanized using a centralized architecture.

Over the past three years, important insights have been gained and significant progress has been made on certain basic issues associated with the development of decentralized fault detection and identification, and control algorithms for non-classical information patterns. These efforts allowed an appreciation of the complexity of the decentralized problem, but more importantly they showed the directions that should be taken to make significant progress in the fundamental issues of distributed estimation, analytical redundancy management, and control.

In particular, control issues involving the data transmission through noisy channels are explored. Furthermore, the problem of detecting faults in local agents and the development of a decentralized methodology for distributed redundancy management was addressed and some resolution to these problems obtained. These results have given new direction in the development of a theory for the control of dynamic networks.

14. SUBJECT TERMS

control of dynamic networks, control with communication constraints, decentralized control, analytic redundancy management, detection filters

15. NUMBER OF PAGES
79

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500 Computer Generated STANDARD FORM 298 (Rev 2-89)
Prescribed by ANSI Std 239-18
298-102

20001208 046

DTIC QUALITY INSPECTED 4

A SYSTEM THEORETIC APPROACH TO AUTONOMOUS VEHICLE DYNAMIC FORMATION

Final Report
Air Force Office of Scientific Research
F49620-97-1-0272
Period of Performance: 04/01/1997 — 12/31/1999

Jason L. Speyer, Principal Investigator
Department of Mechanical and Aerospace Engineering
University of California, Los Angeles

A dynamic network of cooperating agents is characterized by a spatially distributed set of dynamic nodes or agents which coordinate to perform the mission objectives. These coordinated clusters could be composed of an array of satellites constructing a large aperture radar or a swarm of UAV's used to suppress enemy air defenses. These mission objectives are to be achieved in the presence of large uncertainties due largely to a hostile environment. Within this context, nodes may fail at various levels, measurements may be highly corrupted and communication channels may be severely limited due to jamming. Communication links are further challenged due to power constraints and large spatial dispersion, producing tradeoffs between uncertain information, latency and bandwidth constraints. A decision and allocation process appears computationally intractable, especially if mechanized using a centralized architecture.

Over the past three years, important insights have been gained and significant progress has been made on certain basic issues associated with the development of decentralized fault detection and identification, and control algorithms for non-classical information patterns. These efforts allowed an appreciation of the complexity of the decentralized problem, but more importantly they showed the directions that should be taken to make significant progress in the fundamental issues of distributed estimation, analytical redundancy management, and control.

In particular, control issues involving the data transmission through noisy channels are explored. Furthermore, the problem of detecting faults in local agents and the development of a decentralized methodology for distributed redundancy management was addressed and some resolution to these problems obtained. These results have given new direction in the development of a theory for the control of dynamic networks.

Table of Contents

1 Introduction

One of the most important findings of the Air Force Scientific Advisory Board Summer Study on UAV's, from the mission system viewpoint, is that in most operational tasks, UAV's frequently should be employed in coordinated clusters rather than as independent platforms. This notion should also be applied to the use of clusters of satellites configured to produce, for example, a large aperture radar. These clusters of cooperating agents may be characterized by a spatially distributed set of dynamic nodes where individual agents have access to regional information and can share data through a communications network. Within this context, nodes may fail at various levels, measurements may be highly corrupted and communication channels are challenged due to power constraints, noisy information, latency and bandwidth constraints.

The system requirements are induced by the mission objectives, the mission environment, and the system capability. The system capability depends upon the type and number of assets. These assets are fused together by an information system which integrates all available information over all assets or nodes in a wireless communication data network. Upon this data network is imposed a decision system which directs assets such that mission objectives are met. The development of such a system is a great intellectual challenge. Current approaches impose heuristic management architectures on a hierarchy of system functions and use nonparametric schemes to produce the decision processes. From these approaches, little can be understood about the value of information and the decision processes that use this information. Since transmission of information is necessarily limited, it is important to communicate only the information that is most valuable for mission success within the data network. Furthermore, decision rules are to be constructed that best utilize information for the control and guidance of a particular node, or to enhance the awareness of the other nodes with minimal transmissions.

Since the control of dynamic networks is essentially in its infancy, creating general procedures that can be efficiently implemented is a long term goal of any well conceived research program. However, near term goals could be posed which represent important elements of the more complete problem. To this end, over the last three years, significant progress has been made on the development of fault detection filters, decentralized fault detection and identification, and control of systems with nonclassical information patterns. In Section 2, this work is briefly described with supporting documentation in the appendices. In Section 2.1, the effect of noisy information and the type of information that must be transmitted to ensure stable and well performing decentralized control is illuminated. In Section 2.2, the structure of the decentralized detection filter required for analytic redundancy management of a cluster of agents has indicated the direction for a more complete theory on the decomposition of the global system into local systems. For example, this decomposition should achieve the minimal distribution of information in the dynamic network. In a distributed sensing and actuation architecture, individual agents have access to regional information which, perhaps, can be shared under a data communication network. The results of Section 2 are important because they show the directions for developing a systematic methodology for coordinating a distributed set of local systems for global operation.

2 Progress Over the Grant Period

A consequence of cooperative missions is that the agents may require robust, high-performance data networks for information exchange. This need drives a new research direction in the theory of the control of dynamic networks. The essence of determining the architecture for a cooperative system of agents under large uncertainties is to ensure valued information is distributed in a timely way. In Section 2.1 control issues involving the data transmission through noisy channels are explored. It is shown that if the noise is not excessive, then linear controllers with modified gains are nearly optimal. Furthermore, the form of the data to be transmitted has an enormous impact on the system stability as well as performance. In Section 2.2 the problem of detecting faults in the local agents and the development of a decentralized methodology for distributed redundancy management are described. Two important new results are obtained. First, we develop a new robust detection filter based on a disturbance attenuation methodology, which allows the detection and identification of multiple faults. Second, a decentralized detection filter is developed.

2.1 Decentralized Control with Noisy Transmission of Information

In [1, Appendix A], a simple example of a decentralized control problem with noisy information transmission was investigated. This example is a reformulation of the Witsenhausen counterexample which allows the first station to send its information to the second station through an additive white Gaussian noise channel. We show that Witsenhausen's original counterexample can be seen as a limit case in this new formulation. We believe that this new formulation is closer to many applications in large scale systems, where different pieces of information could be transmitted among the stations through some noisy channels. We should note that as soon as some kind of communication uncertainty is introduced for the transmission, the information pattern is no longer classical and the cost may no longer be convex in the strategies. Hence, the optimal strategies, which may not even be unique, are very difficult to find. Similar approaches that have so far been used for the Witsenhausen problem, might be applied to this new formulation as well. For example, asymptotic approaches using expansions in small ϵ can be used.

In [2, Appendix B], we considered the case where the communication uncertainty is small. We followed an asymptotic approach where we approximated the cost based on its expansion in terms of the small transmission noise intensity. We showed how minimizing the approximated cost can be seen as a singular optimization problem. We then used a variational approach in order to find the necessary conditions for the asymptotically optimal strategies and showed that some reasonable linear strategies would actually satisfy those conditions. We also provided some intuitive explanations for the behavior of those linear strategies and obtained their corresponding cost. All the derivations and results in this paper show some of the difficulties involved in dealing with decentralized systems as soon as we deviate a little bit from a classical, or at least a partially nested, information pattern. On the other hand, although we have modeled the communication uncertainty in the simplest possible way, we have tried to emphasize the role of communication uncertainties in generating information patterns that are very difficult to handle. Even though the optimization problem is generally difficult for this class of systems, in some applications we might be able to exploit the specific structure of the system in order to obtain some reasonably good sup-optimal strategies, which would yield an acceptable performance.

Finally, in [3, Appendix C], a more general two-station decentralized LQG problem was formulated, where the local controllers had to be designed based on some local information in order to minimize a single common cost. This problem generally has a non-classical information pattern and the optimal controls are usually unknown. One of the first possible sub-optimal approaches is to decompose the problem into separate centralized problems. In this paper, we investigated such an approach for different communication scenarios between the stations, namely, when the stations communicate their controls, their measurements or both, or their estimation residuals. We showed that even though our approach is quite reasonable for the case where the stations communicate all their measurements, it may fail to stabilize the closed-loop system as soon as the compensator is unstable. Then, we showed how this difficulty can be removed if the stations either communicate both their measurements and their controls or communicate their estimation residuals. All these results show some of the fundamental differences between the centralized and the decentralized structures. Moreover, we have tried to elaborate on the role of communication among the stations and corresponding uncertainties. While many new applications for spatially distributed dynamic systems are emerging, there are still major difficulties that need to be addressed.

2.2 Detection Filters for Robust Analytical Redundancy

Any system under automatic control demands a high degree of system reliability. Therefore, the system relies on the health of the sensors, plant, and actuators. If a system fault occurs, the controller will not work properly. If a sensor fails, the command generated by the controller will be based on the wrong information. If an actuator fails, the controller's command will not be executed properly in the system. Therefore, a health monitoring system capable of detecting a fault as it occurs and identifying the faulty component is required. The most common approach is hardware redundancy, which is the direct comparison of identical components. This approach requires very little computation. However, hardware redundancy is expensive and limited by space and weight. An alternative is analytical redundancy, which uses a modeled dynamic relationship between system inputs and measured system outputs to form a residual process used for detecting and identifying faults. Nominally, the residual is nonzero only when a fault has occurred and is zero at other times. Therefore, no redundant components are needed. However, additional computation is required.

A popular approach to analytical redundancy is the detection filter which was first introduced by [4] and refined by [5]. It is also known as the Beard-Jones fault detection filter. A geometric interpretation of this filter is given in [6] and a spectral theory and implementation appeared in [7]. Design algorithms have been developed [8,9] which improved detection filter robustness. The idea of a detection filter is to put the reachable subspace of each fault into invariant subspaces which do not overlap with each other. Then, when a nonzero residual is detected, a fault can be announced and identified by projecting the residual onto each of the invariant subspaces. Therefore, multiple faults can be monitored in one filter.

Another related approach, the unknown input observer [10], simplifies the detection filter problem by dividing the faults into a target fault and a group of nuisance faults where the nuisance faults are placed into one unobservable subspace. Although only one fault can be detected in each unknown input observer, the additional flexibility in robust fault detection filter design for general time-varying systems is obtained by using this approximate fault detection filter.

Four fault detection and identification algorithms were developed, progressively improving the relationship between robustness and detection and identification. They are the game theoretic fault detection filter (an approximate unknown input observer), the optimal stochastic fault detection filter (an approximate unknown input observer), the residual-sensitive fault detection filter (an approximate unknown input observer), and the optimal stochastic multiple-fault detection filter (an approximate Beard-Jones fault detection filter).

2.2.1 A Game-Theoretic Fault Detection Filter

In [11] we posed and solved a disturbance attenuation problem which closely approximates the actions of a fault detection filter. The end product is a game theoretic filter which acts as an approximate unknown input observer. We also showed that this approximation can be made more and more exact until, in the limit, the game theoretic filter becomes an unknown input observer exactly. A related result is that a reduced-order observer can also be obtained from the limiting case. The disturbance attenuation-based approach that we have introduced here leads to filters which are more flexible, more robust, and more applicable than existing fault detection structures. This approach allows time-varying systems to be monitored for the first time. Finally, in the course of our limiting case analysis, we showed that singular optimization theory can be used to analyze the asymptotic properties of game theoretic estimators. It is possible that the application of singular optimization theory to other disturbance attenuation problems can lead to similar insights.

2.2.2 Optimal Stochastic Fault Detection Filter

Properties of the optimal stochastic fault detection filter for fault detection and identification are determined in [12, Appendix C]. The objective of the filter is to monitor a certain fault called the target fault and block other faults which are called nuisance faults. This filter is derived by keeping the ratio of the transmission from nuisance faults to the transmission from the target fault small. Rather than an arbitrary function, the fault amplitudes are modeled as white noise input processes. It is shown that this filter approximates the properties of the classical fault detection filter such that in the limit, where the ratio of the transmissions is zero, the optimal stochastic fault detection filter is equivalent to the unknown input observer. However, the nuisance fault directions and their associated invariant zero directions must be included in the invariant subspace generated by this fault detection filter. Fault detection filter designs can be obtained for both linear time-invariant and time-varying systems.

2.2.3 A Generalized Least-Squares Fault Detection Filter

The generalized least-squares fault detection filter in [13, Appendix D] is derived from solving a min-max problem which makes the residual sensitive to the target fault, but not to the nuisance fault. This is an alternate derivation of the optimal stochastic fault detection filter [12] of Section 2.2.2. In the limit, as the nuisance fault weighting goes to zero, this filter is equivalent to an unknown input observer which puts the nuisance fault into an unobservability subspace. Furthermore, there exists a reduced-order filter in the limit. Since the target fault is explicit in this derivation, the reduced-order filter is found witb respect to the target fault direction and weighting. This aspect is different from that of the game theoretic detection filter [11] where this dependence does not exist. This filter also extends the unknown input observer to a time-varying system.

2.2.4 Optimal Stochastic Multiple-Fault Detection Filter

The optimal stochastic multi-fault detection filter [14, Appendix E] is a generalization of the optimal stochastic single-fault filter. The residual space of the filter is divided into several subspaces and each subspace is sensitive to only its target fault, but not the nuisance faults, in the sense that the ratio of the transmission from the nuisance faults to the transmission from target fault is small. In the limit as the ratio goes to zero and in the absence of sensor noise and a complementary subspace, this filter is equivalent to a Beard-Jones fault detection filter which puts each fault into an unobservable subspace. This filter has the advantages of the unknown input observer in that it can be designed for robustness and time-varying systems, and the advantages of the Beard-Jones fault detection filter by being capable of detecting multiple faults in one filter. Although there is additional computation to determine the filter gain and projectors, this can be done off-line so that implementation is as straightforward as the Beard-Jones fault detection filter.

2.2.5 A Decentralized Fault Detection Filter

In [15, Appendix F] we introduce the decentralized fault detection filter which is the structure that results from merging decentralized estimation theory with the game theoretic fault detection filter. A decentralized approach may be the ideal way to monitor the health of large-scale systems since it breaks the problem down into smaller pieces and it is easily scalable. An essential feature is that the local measurements, which may include the information of the relative state space between agents, and the fault direction, which may also be associated with the inter agent measurements, produce local state spaces from the global state by a minimal realization. This local state space contains information associated with multiple agents.

3 References

1. K. Shoarinejad, J.L. Speyer, and I. Kanellakopoulos, "Decentralized Control with Noisy Transmission of Information," Proc. American Control Conference, June 1999.

2. K. Shoarinejad, J.L. Speyer, and I. Kanellakopoulos, "An Asymptotic Optimal Design for a Decentralized System with Noisy Communication," Proc. IEEE Conference on Decision and Control, Dec. 1999.

3. K. Shoarinejad, J.D. Wolfe, J.L. Speyer, and I. Kanellakopoulos, "A Two-Station Decentrlized LQG Problem with Noisy Communication," Proc. IEEE Conference on Decision and Control, Dec. 1999.

4. R.V. Beard, *Failure Accomodation in Linear Systems through Self-Reorganization*, Ph.D. Thesis, Massachusetts Institute of Technology, 1971.

5. H.L. Jones, *Failure Detection in Linear Systems*, Ph.D. Thesis, Massachusette Institute of Technology, 1973.

6. M.-A. Massoumnia, "A Geometric Approach to the Synthesis of Failure Detection Filters," *IEEE Trans. Automatic Control*, **31(9)**:839-846, 1986.

7. J.E. White and J.L. Speyer, "Detection Filter Design: Spectral Theory and Algorithms," *IEEE Trans. Automatic Control*, **32(7)**:593-603, 1987.

8. R.K. Douglas and J.L. Speyer, "Robust Fault Detction Filter Design," *AIAA J. Guidance, Control, and Dynamics*, **19(1)**:214-218, 1996.

9. R.K. Douglas and J.L. Speyer, "$\mathcal{H}_\infty$ Bounded Fault Detection Filter," *AIAA J. Guidance, Control, and Dynamics*, **22(1)**:129-138, 1999.

10. M.-A. Massoumnia, G.C. Verghese, and A.S. Willsky, "Failure Detection and Identification," *IEEE Trans. Automatic Control*, **34(3)**:316-321, 1989.

11. W.H. Chung and J.L. Speyer, "A Game Theoretic Fault Detection Filter," *IEEE Trans. Automatic Control*, **43(2)**:143-161, 1998.

12. R.H. Chen and J.L. Speyer, "Optimal Stochastic Fault Detection Filter," Proc. American Control Conf., 1999.

13. R.H. Chen and J.L. Speyer, "A Generalized Least-Squares Fault Detection Filter," *J. Adaptive Control Signal Process.*, **14**:747-757, 2000.

14. R.H. Chen and J.L. Speyer, "Optimal Stochastic Multiple-Fault Detection Filter," Proc. 38^{th} Conf. on Decision and Control, 1999.

15. W.H. Chung and J.L. Speyer, "A Decentralized Fault Detection Filter," Proc. American Controls Conf., June 1999; also submitted for publication to ASME, 1999.

APPENDIX A

Decentralized Control with Noisy Transmission of Information

K. Shoarinejad, J.L. Speyer, I. Kanellakopoulos

Decentralized Control with Noisy Transmission of Information *

Kambiz Shoarinejad
UCLA Electrical Engineering
Box 951594
Los Angeles, CA 90095-1594
Tel: (310) 206-1022
Fax: (310) 206-1022
e-mail: kambiz@ee.ucla.edu

Jason L. Speyer
UCLA Mechanical and Aerospace Engineering
Box 951597
Los Angeles, CA 90095-1597
Tel: (310) 206-4451
Fax: (310) 206-2302
e-mail: speyer@seas.ucla.edu

Ioannis Kanellakopoulos
UCLA Electrical Engineering
Box 951594
Los Angeles, CA 90095-1594
Tel: 310-206-5278
Fax: 310-206-1022
e-mail: ioannis@ee.ucla.edu

Abstract

We consider an example for a decentralized stochastic optimal control problem, where the non-classical nature of the information pattern is induced by the transmission noise in the system. This example is a reformulation of the Witsenhausen counterexample, where the first station is allowed to send its information to the second station through an additive white Gaussian noise channel. We establish the non-convexity of the problem in this new formulation and show that the problem considered here converges asymptotically to the classical problem or the Witsenhausen problem, as the transmission noise intensity converges to zero or diverges to infinity, respectively.

1 Introduction

Dealing with large scale systems has become a great challenge for systems analysts and engineers more than ever. There are many such systems, which are composed of a large number of complex interconnected subsystems and hence do not satisfy the centrality assumption that is prevalent among classical engineering approaches. One of the main characteristics of these systems is that distributed decisions must be made based on decentralized information. Different stations may communicate with each other, possibly by signaling through noisy channels. The control problem is to develop coordinated strategies for the stations in order to achieve a common objective.

The way that information is distributed in a decentralized system highly affects the performance of the controlled system. Changes in the information pattern will produce changes in the optimal achievable cost. Even though there are always some constraints on how the information can be distributed in a physical system (where to put the sensors and the actuators, what to transmit, etc.), in general, there are many possible information patterns for a given system.

When the stations do not have access to the same information and/or some stations do not have perfect recall, i.e., they lose information, we have a non-classical information pattern. Optimal strategies for decentralized systems with general non-classical patterns are still unknown. One main difficulty is that the information available to one station may not be sufficient to determine the previous actions by other stations, which have affected that information. This will destroy the convexity of the cost function with respect to the strategies, even though it may look convex in the controls.

In 1968, Witsenhausen provided a simple example in [8], where there are only two stations, the dynamics are linear, the underlying uncertainties are additive and Gaussian and the cost is quadratic. The information pattern, however, is non-classical. He established the existence of the optimal design and by proposing a nonlinear set of strategies, showed that no affine strategy could be optimal. This seemingly simple example, which is also called Witsenhausen's counterexample, turned out to be extremely hard. It is still outstanding after 30 years. This example in fact motivated much research on the links between decentralized stochastic control problems and team theory and the effects of different information patterns on decentralized systems. Although it is a very simple example, it demonstrates the main difficulties induced by non-classical information patterns.

In the next section, we reformulate Wisenhausen's problem by assuming that the first station sends its information to the second station through a noisy channel. In Section 3, we obtain an alternative form for the performance index in this new formulation, which shows the possible non-convexity of the cost with respect to the strategies. In Section 4, we consider two limit cases, namely when the transmission noise intensity is small and when it becomes very large. We will see how this new formulation covers a wide range of problems from classical LQG problem to the Witsenhausen counterexample. Finally, Section 5 contains some concluding remarks.

* This research was supported in part by the National Science Foundation under Grant ECS-9502945, Air Force Office of Scientific Research under Grant F49620-97-1-0272 and Office of Naval Research under Award N00014-97-1-0939

2 Problem Statement

Consider a two-stage stochastic problem with the following state equations:

$$\begin{aligned} x_1 &= x_0 + u_1 & (2.1) \\ x_2 &= x_1 - u_2, & (2.2) \end{aligned}$$

where x_0 is the random initial state, which is assumed to be Gaussian with zero mean and variance σ_0^2. The information available to the two stations is determined by the following output equations:

$$\begin{aligned} z_1 &= x_0 & (2.3) \\ z_2 &= \begin{bmatrix} x_0 + \epsilon v_t \\ x_0 + u_1 + v_2 \end{bmatrix} \triangleq \begin{bmatrix} z_{21} \\ z_{22} \end{bmatrix}, & (2.4) \end{aligned}$$

where v_2 is the measurement noise for the second station, which is again assumed to be a zero mean Gaussian random variable with unit variance. As we can see, the information available to the first station is being transmitted to the second station through an additive white Gaussian noise channel with $\epsilon v_t \sim \mathcal{N}\left(0, \epsilon^2\right)$ being the transmission noise. Also x_0, v_2 and v_t are all assumed to be independent of each other. Note that the communication uncertainty is simply modeled as an additive Gaussian noise. This model may not be very realistic when digital communication is used. However, since there are already some major difficulties in dealing with non-classical information patterns, using more complicated models for the communication uncertainties may not be very reasonable at this point.

The objective is now to design the controls:

$$\begin{aligned} u_1 &= \gamma_1\left(z_1\right) & (2.5) \\ u_2 &= \gamma_2\left(z_2\right), & (2.6) \end{aligned}$$

in order to minimize the following cost function:

$$J = E\left[k^2 u_1^2 + x_2^2\right], \tag{2.7}$$

where $k^2 > 0$ is a given constant. We see that the first controller has perfect information but its action is costly. In contrast, the second controller has inexpensive control but noisy information. Since the second station does not know exactly what the first station knew, due to the communication uncertainty, we don't have perfect recall and hence we still have a non-classical pattern. If there was no transmission noise, we would have a classical information pattern for which a set of strategies, which are linear in the information, is known to be the unique optimal solution.

3 An Alternative Form for the Performance Index

In this section, the performance index is rewritten in terms of the Fisher information matrix, which indicates that the cost may not be convex in the strategies.

Similar to the Witsenhausen problem, we define:

$$\begin{aligned} f\left(z_1\right) &\triangleq z_1 + \gamma_1\left(z_1\right) = x_0 + u_1 & (3.1) \\ g\left(z_2\right) &\triangleq \gamma_2\left(z_2\right) = u_2. & (3.2) \end{aligned}$$

Then the cost can be expressed as:

$$\begin{aligned} J &= E\left[k^2 u_1^2 + x_2^2\right] \\ &= E\left[k^2\left(z_1 - f\left(z_1\right)\right)^2 + \left(f\left(z_1\right) - g\left(z_2\right)\right)^2\right] \\ &\triangleq J\left(f, g\right). & (3.3) \end{aligned}$$

It is clear that for a fixed strategy f, the optimal strategy g is the conditional expectation, i.e.,:

$$g^*\left(z_2\right) = \arg\min_g J\left(f, g\right) = E\left[f\left(z_1\right) | z_2\right]. \tag{3.4}$$

Substituting back in the cost, we get:

$$\begin{aligned} J^*(f) &\triangleq J\left(f, g^*\right) \\ &= k^2 E\left[\left(z_1 - f\left(z_1\right)\right)^2\right] + E\left[\left(f\left(z_1\right) - g^*\left(z_2\right)\right)^2\right] \\ &= k^2 E\left[\left(z_1 - f\left(z_1\right)\right)^2\right] + E\left[\left(f\left(z_1\right)\right)^2\right] - E\left[\left(g^*\left(z_2\right)\right)^2\right] & (3.5) \end{aligned}$$

where we have used the orthogonality property of the conditional expectation:

$$E\left[\left(f\left(z_1\right) - g^*\left(z_2\right)\right) g^*\left(z_2\right)\right] = 0. \tag{3.6}$$

It is important to note the minus sign in the third term in (3.5). As we shall see, this minus sign could indeed destroy the convexity of the cost with respect to the strategies.

On the other hand:

$$\begin{aligned} g^*\left(z_2\right) &= \int f\left(z_1\right) p\left(z_1 | z_2\right) dz_1 \\ &= \frac{\int f\left(z_1\right) p\left(z_1, z_2\right) dz_1}{\int p\left(z_1, z_2\right) dz_1}, & (3.7) \end{aligned}$$

where $p\left(z_1, z_2\right)$ is the joint probability density of z_1 and z_2. The following lemma can be used in order to express $g^*\left(z_2\right)$ in terms of z_2 and its probability density.

Lemma 3.1: $f\left(z_1\right) p\left(z_1, z_2\right)$ can be expressed in terms of z_{22} and the joint probability density of z_1 and z_2 in the following form:

$$f\left(z_1\right) p\left(z_1, z_2\right) = z_{22} p\left(z_1, z_2\right) + \frac{\partial}{\partial z_{22}} p\left(z_1, z_2\right). \tag{3.8}$$

Proof:

$$\begin{aligned} & z_{22} p\left(z_1, z_2\right) + \frac{\partial}{\partial z_{22}} p\left(z_1, z_2\right) \\ &= z_{22} p\left(z_1, z_2\right) + \frac{\partial}{\partial z_{22}} p\left(z_2 | z_1\right) p\left(z_1\right) \end{aligned}$$

$$
\begin{aligned}
&= z_{22} p(z_1, z_2) \\
&+ \frac{\partial}{\partial z_{22}} p_{(\epsilon v_t, v_2)} \left(\begin{bmatrix} z_{21} \\ z_{22} \end{bmatrix} - \begin{bmatrix} z_1 \\ f(z_1) \end{bmatrix} \right) p(z_1) \\
&= z_{22} p(z_1, z_2) \\
&+ \frac{\partial}{\partial z_{22}} \left(\frac{1}{2\pi\epsilon} \exp\left(-\frac{(z_{21} - z_1)^2}{2\epsilon^2} - \frac{(z_{22} - f(z_1))^2}{2} \right) \right) p(z_1) \\
&= f(z_1) p(z_1, z_2), \qquad (3.9)
\end{aligned}
$$

where we have used the specific form of the information available to the second station and the fact that $\epsilon v_t \sim \mathcal{N}(0, \epsilon^2)$ and $v_2 \sim \mathcal{N}(0, 1)$ are independent.

Substituting for $f(z_1) p(z_1, z_2)$ from the above lemma back in (3.7) and integrating with respect to z_1, we will obtain $g^*(z_2)$ as follows:

$$
g^*(z_2) = z_{22} + \frac{\partial}{\partial z_{22}} \ln p(z_2). \qquad (3.10)
$$

On the other hand, we have:

$$
E\left[z_{22}^2\right] = E\left[(f(z_1))^2\right] + 1, \qquad (3.11)
$$

and:

$$
E\left[z_{22} \frac{\partial}{\partial z_{22}} \ln p(z_2)\right] =
$$

$$
\int \int_{-\infty}^{+\infty} z_{22} \frac{\partial}{\partial z_{22}} \ln (p(z_{21}, z_{22})) p(z_{21}, z_{22}) \, dz_{21} dz_{22}. \quad (3.12)
$$

If we integrate by parts with respect to z_{22}, we will get:

$$
\begin{aligned}
&\int_{-\infty}^{+\infty} z_{22} \frac{\partial}{\partial z_{22}} \ln (p(z_{21}, z_{22})) p(z_{21}, z_{22}) \, dz_{22} \\
&= z_{22} p(z_{21}, z_{22}) \big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} p(z_{21}, z_{22}) \, dz_{22} \\
&= -p(z_{21}), \qquad (3.13)
\end{aligned}
$$

where z_{22} is assumed to have a finite mean value and therefore the first term becomes zero. Hence:

$$
E\left[z_{22} \frac{\partial}{\partial z_{22}} \ln p(z_2)\right] = -1. \qquad (3.14)
$$

We can now obtain $E\left[(g^*(z_2))^2\right]$ and substitute it back in (3.5) to express the performance index in the following form:

$$
J^*(f) = k^2 E\left[(z_1 - f(z_1))^2\right] + 1 - I_f(Z_2)_{22}, \quad (3.15)
$$

where $I_f(Z_2)_{22}$ is indeed the $(2,2)$ element of the *Fisher information matrix*[1] for z_2, which is defined as follows:

$$
I_f(Z_2) \triangleq E\left[\nabla_{z_2}^T \ln p(z_2) \cdot \nabla_{z_2} \ln p(z_2)\right]. \qquad (3.16)
$$

[1] Fisher information is originally obtained in the Cramer-Rao bound, which is a measure for the minimum error in estimating a parameter based on the value of a random variable. However, by introducing a location parameter, an alternative form of the Fisher information may be defined for a random variable with a given distribution. This alternative form is in fact related to the entropy measure (see [4], p.494).

The subscript f indicates the fact that it actually depends on the form of the strategy f, which is present in the definition of z_2 and would affect its probability density function. As we see, the cost is now expressed only in terms of one strategy f. Also, this somehow shows us that in order to minimize the cost, we need to get the lowest possible cost associated with the first station, while we transfer as much information as possible to the second station through the dynamics of the system. The possible non-convexity of the cost with respect to f can also be seen from the above expression. It can be shown that the Fisher information term is a convex functional [3]. Therefore, $1 - I_f(Z_2)_{22}$ is concave and the sum of a convex and a concave functional may not be convex.

4 Limit Cases

4.1 Noiseless Transmission

We first consider the limit case in which the transmission is noiseless, i.e., $\epsilon = 0$ and hence $z_{21} = z_1$. In this case, the second station knows exactly what the first station knew. Therefore, we have perfect recall and the information pattern is classical. We can write:

$$
\begin{aligned}
p(z_2) &= p(z_{21}, z_{22}) = p(z_{22} | z_{21}) p(z_{21}) = p(z_{22} | z_1) p(z_1) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z_{22} - f(z_1))^2}{2} \right) p(z_1). \qquad (4.1)
\end{aligned}
$$

Then, from (3.10), we will have:

$$
g^*(z_2) = f(z_1) = f(z_{21}), \qquad (4.2)
$$

which could easily be obtained from the original definition for g^*, i.e.,:

$$
g^*(z_2) = E[f(z_1) | z_2] = f(z_1), \qquad (4.3)
$$

because z_1 is exactly known given z_2. Substituting this back in (3.5) and minimizing with respect to the strategy f, we will have:

$$
g^*(z_2) = f(z_1) = z_1, \qquad (4.4)
$$

and hence:

$$
\begin{aligned}
\gamma_1(z_1) &= 0 \qquad (4.5) \\
\gamma_2(z_2) &= z_1, \qquad (4.6)
\end{aligned}
$$

which is the unique linear set of optimal strategies. This indeed turns out to be a very simple example of the well-known LQG problems.

4.2 Infinite Transmission Noise Intensity

Another limit case is when the transmission noise intensity increases to infinity. In this case, z_{21} and z_{22} indeed become independent and we will have:

$$
p(z_2) = p(z_{21}, z_{22}) = p(z_{21}) p(z_{22}). \qquad (4.7)
$$

The Fisher information term can now be written as:

$$\begin{aligned} & I_f(Z_2)_{22} = \\ & \int\int_{-\infty}^{+\infty}\left(\frac{\partial}{\partial z_{22}}\ln p(z_{21},z_{22})\right)^2 p(z_{21},z_{22})\,dz_{21}dz_{22} \\ &= \int_{-\infty}^{+\infty}\left(\frac{\partial}{\partial z_{22}}\ln p(z_{22})\right)^2 p(z_{22})\,dz_{22} \\ &= I_f(Z_{22}), \end{aligned} \tag{4.8}$$

which is actually the Fisher information content of z_{22} only. Hence:

$$J^*(f) = k^2 E\left[(z_1 - f(z_1))^2\right] + 1 - I_f(Z_{22}). \tag{4.9}$$

This is the same result that was presented for the Witsenhausen counterexample in [8]. Intuitively, when we have infinite transmission noise, we might as well deny the access to z_1 for the second station, which is exactly the case in Witsenhausen's counterexample. The optimal strategies for this case are still unknown. Witsenhausen showed that the optimal solution exists, even if x_0 has a general distribution with a finite second moment [8]. He then showed that if one of the strategies is restricted to be affine, the other optimal strategy would also be affine. But then he provided a set of nonlinear strategies that could achieve a lower cost for some values of k^2 and σ_0. Different approaches have been taken in order to find the optimal strategies. The asymptotic approach was used in [2] for the case where σ_0 is small. In [1], a neural network, trained by stochastic approximation techniques, was used in order to approximate the optimal strategies. It was demonstrated that the optimal $f^*(z_1)$ may not be strictly piecewise, as was suggested by Witsenhausen, but slightly sloped. Some researchers have tried to attack the problem numerically and use some sample and search techniques to find the solution. A discretized version of the problem was formulated in [5], which was later shown in [7] to be NP-complete and computationally intractable. It is recently asserted in [6] that a global optimum would be achieved by searching directly in the strategy space using the generalized step functions to approximate $f(z_1)$.

5 Concluding Remarks

A simple example of a decentralized control problem with noisy information transmission was investigated. This example is a reformulation of the Witsenhausen counterexample by allowing the first station to send its information to the second station through an additive white Gaussian noise channel. In fact, we show that Witsenhausen's original counterexample can be seen as a limit case in this new formulation. We believe that this new formulation is closer to many applications in large scale systems, where different pieces of information could be transmitted among the stations through some noisy channels. We should note that as soon as some kind of communication uncertainty is introduced for the transmission, the information pattern is no longer classical and the cost may no longer be convex in the strategies. Hence, the optimal strategies, which may not even be unique, are very difficult to find. Similar approaches that have so far been used for the Witsenhausen problem, might be applied to this new formulation as well. For example, asymptotic approaches using expansions in small ϵ are possible.

References

[1] Baglietto,M. and Parisini, T. and Zoppoli, R., "Nonlinear approximations for the solution of team optimal control problems", *Proceedings of CDC 1997*, Vol. 5, pp. 4592-4594, 1997.

[2] Castanon, D. A. and Sandell,Jr., N. R., "Signaling and uncertainty: a case study", *Proceedings of CDC 1978*, pp. 1140-1144, 1978.

[3] Cohen, M. L., "The Fisher information and convexity", *IEEE Trans. on Information Theory*, Vol. 14, pp. 591-592, 1968.

[4] Cover, T. M. and Thomas, J. A., *Elements of information theory*, New York:Wiley, 1991.

[5] Ho, Y. C. and Chang, T. S., "Another look at the nonclassical information structure problem", *IEEE Trans. on Automatic Control*, Vol. AC-25, pp. 537-540, 1980.

[6] Lee, J. T. and Lau, E. and Ho, Y. C., "On the global optimum of the Witsenhausen counterexample", *Draft Version*, 1998.

[7] Papadimitriou, C. H. and Tsitsiklis, J. N., "Intractable problems in control theory", *SIAM Journal on Control and Optimization*, Vol. 24, pp. 639-654, 1986.

[8] Witsenhausen, H.S., "A counterexample in stochastic optimal control", *SIAM Journal of Control*, Vol. 6, pp. 131-147, 1968.

APPENDIX B

An Asymptotic Optimal Design for a Decentralized System with Noisy Communication

K. Shoarinejad, J.L. Speyer, I. Kanellakopoulos

Proceedings of the 38th
Conference on Decision & Control
Phoenix, Arizona USA • December 1999

An Asymptotic Optimal Design for a Decentralized System with Noisy Communication*

Kambiz Shoarinejad
UCLA Electrical Engineering
Box 951594
Los Angeles, CA 90095-1594
Tel: (310) 206-1022
Fax: (310) 206-1022
e-mail: kambiz@ee.ucla.edu

Jason L. Speyer
UCLA Mechanical and Aerospace Engineering
Box 951597
Los Angeles, CA 90095-1597
Tel: (310) 206-4451
Fax: (310) 206-2302
e-mail: speyer@seas.ucla.edu

Ioannis Kanellakopoulos
UCLA Electrical Engineering
Box 951594
Los Angeles, CA 90095-1594
Tel: 310-206-5278
Fax: 310-206-1022
e-mail: ioannis@ee.ucla.edu

Abstract

A reformulation of the Witsenhausen counter-example is considered, where the first station is allowed to transmit its information to the second station through a low noise channel. This is in fact a decentralized stochastic system where the communication uncertainty induces a non-classical information pattern. Assuming a small transmission noise intensity, an asymptotic approach is used in order to find an approximated cost. Then, the necessary conditions for asymptotically optimal strategies are obtained using a variational approach. It is shown that the necessary conditions are satisfied by linear strategies with slightly different coefficients than the noiseless transmission case.

1 Introduction

Coordinating and controlling dynamic systems in spatial networks has always been a challenging problem for system designers. It is now attracting more attention as various new applications are emerging in a very wide range, from controlling autonomous vehicles in formation to flow and congestion control in computer networks. However, there are still some major difficulties in dealing with such systems.

The main characteristics of any decentralized system is that the information is distributed among different stations and the performance of the system highly depends on the corresponding information pattern, i.e., who knows what and when. The stations may communicate with each other, possibly by signaling through noisy channels. Even though there might be some physical constraints on the information structure of the system (e.g. locations of the sensors, the actuators, and the transmitters), in general, an optimal information pattern should be obtained. Then, based on the locally available information, a set of coordinated local strategies should be designed in order to achieve a common objective. In many cases, however, we will end up with non-convex functional optimization problems, which are usually very difficult to solve.

One such class of problems is when the decentralized system has a non-classical information pattern which is not partially nested. In this case, some stations can not reconstruct the previous actions of other stations which have affected their own local information. Unfortunately, this happens in many decentralized systems.

In 1968, Witsenhausen provided a simple example in [6], where there are only two stations, the dynamics are linear, the underlying uncertainties are additive and Gaussian and the cost is quadratic. The information pattern, however, is non-classical. This example, which demonstrates some of the major difficulties in dealing with non-classical information patterns, motivated much research on the links between decentralized stochastic control problems and team theory and the effects of different information patterns on decentralized systems.

In this example, one station acts first and affects the information available to the next station while there is no way for the second station to determine the action of the first station. The existence of the optimal design was established in [6], where a nonlinear set of strategies was also proposed which showed that no affine strategy could be optimal. It was later shown in [3] that when the uncertainty on the information available to the first station is small, linear strategies would still be optimal over a large class of nonlinear strategies. Intuitively, when the uncertainty on the information of the first station is small, the second station will also be able to *guess* what that information was. Therefore, since the problem is cooperative in the sense that the stations are aware of each others' strategies, the second station can almost reconstruct the action of the first station and there is no need for any kind of signaling among the stations through the dynamics of the system.

*This research was supported in part by the National Science Foundation under Grant ECS-9502945, Air Force Office of Scientific Research under Grant F49620-97-1-0272 and Office of Naval Research under Award N00014-97-1-0939

$$g(z_2) \triangleq \gamma_2(z_2) = u_2. \tag{2.9}$$

Then the cost can be expressed as:

$$\begin{aligned} J &= E\left[k^2 u_1^2 + x_2^2\right] \\ &= E\left[k^2 (z_1 - f(z_1))^2 + (f(z_1) - g(z_2))^2\right] \\ &\triangleq J(f,g). \end{aligned} \tag{2.10}$$

If we fix the function f, the optimal strategy g will clearly be obtained as the conditional expectation, i.e.,:

$$g^*(z_2) = \arg\min_g J(f,g) = E[f(z_1)|z_2]. \tag{2.11}$$

It was shown in [5] that:

$$g^*(z_2) = z_{22} + \frac{\partial}{\partial z_{22}} \ln p(z_2). \tag{2.12}$$

where $p(z_2) = p(z_{21}, z_{22})$ is the probability density function for the information available to the second station. It was further shown that the cost can be written as the following and may not be convex in f:

$$\begin{aligned} J^*(f) &\triangleq J(f,g^*) \\ &= k^2 E\left[(z_1 - f(z_1))^2\right] + E\left[(f(z_1) - g^*(z_2))^2\right] \\ &= k^2 E\left[(z_1 - f(z_1))^2\right] + E\left[(f(z_1))^2\right] - E\left[(g^*(z_2))^2\right] \quad (2.13) \\ &= k^2 E\left[(z_1 - f(z_1))^2\right] + 1 - I_f(Z_2)_{22}, \quad (2.14) \end{aligned}$$

where $I_f(Z_2)_{22}$ is the $(2,2)$ element of the *Fisher information matrix* for z_2, which is defined as:

$$I_f(Z_2) \triangleq E\left[\nabla_{z_2}^T \ln p(z_2) \cdot \nabla_{z_2} \ln p(z_2)\right]. \tag{2.15}$$

As we mentioned earlier, for the noiseless transmission case, the unique optimal strategies, which are linear in the information, are easily obtained. On the other hand, when the transmission noise intensity ϵ is small, we would still expect a similar behavior for the optimal strategies. In the following sections, we will consider this case. Namely, we will assume v_t has a small intensity. Under this assumption, we will obtain the first few terms in the expansion of the cost in terms of ϵ. We will then use the Hamiltonian approach in order to find the necessary conditions for the strategies that minimize the approximated cost.

3 An Expansion for the Cost

Assume that the first station communicates with the second station through a low noise channel. In other words, the transmission noise intensity ϵ is assumed to be small. In this section, we will find an expansion for the cost in terms of ϵ. For this purpose, we first find an expansion for the probability density function of the information available to the second station, i.e., $p(z_2)$. Then, we use (2.12) in order to find the corresponding expansion for $g^*(z_2)$. By substituting back in (2.13), we will obtain the expanded cost only in terms of f.

The probability density function for z_2 can be written as follows:

$$p_\epsilon(z_2) \triangleq p(z_2) = \int_{-\infty}^{+\infty} p(z_{22}, z_{21}, z_1)\, dz_1 \tag{3.1}$$

$$= \int_{-\infty}^{+\infty} p(z_{22}|z_{21}, z_1)\, p(z_{21}|z_1)\, p(z_1)\, dz_1 \tag{3.2}$$

$$= \int_{-\infty}^{+\infty} p(z_{22}|z_1)\, p(z_{21}|z_1)\, p(z_1)\, dz_1 \tag{3.3}$$

$$= \int_{-\infty}^{+\infty} p(z_{22}|z_1)\, p_{v_t}(z_{21} - z_1)\, p(z_1)\, dz_1 \tag{3.4}$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z_{22} - f(z_1))^2}{2}\right) \frac{1}{\sqrt{2\pi}\epsilon} \exp\left(-\frac{(z_{21} - z_1)^2}{2\epsilon^2}\right) \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{z_1^2}{2\sigma_0^2}\right) dz_1, \tag{3.5}$$

where for (3.3) we have used the facts that the σ-fields generated by $\{z_{21}, z_1\}$ and $\{z_1, v_t\}$ are the same and z_1, v_t and v_2 are mutually independent. For small ϵ, we now approximate $\ln p_\epsilon(z_2)$ by considering only the first three terms of its expansion around $\epsilon = 0$. Namely:

$$\ln p_\epsilon(z_2) \simeq \ln p_0(z_2) + \frac{\partial}{\partial \epsilon} \ln p_\epsilon(z_2)\Big|_{\epsilon=0} \epsilon + \frac{\partial^2}{\partial \epsilon^2} \ln p_\epsilon(z_2)\Big|_{\epsilon=0} \epsilon^2. \tag{3.6}$$

By making the following change of variables:

$$\epsilon y \triangleq z_1 - z_{21} \Rightarrow \epsilon dy = dz_1, \tag{3.7}$$

we can write $p_\epsilon(z_2)$ in the following form:

$$p_\epsilon(z_2) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z_{22} - \bar{f}_\epsilon(y))^2}{2}\right) \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{(z_{21} + \epsilon y)^2}{2\sigma_0^2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy, \tag{3.8}$$

where:

$$\bar{f}_\epsilon(y) \triangleq f(\epsilon y + z_{21}). \tag{3.9}$$

It is now clear that:

$$p_0(z_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z_{22} - f(z_{21}))^2}{2}\right) \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{z_{21}^2}{2\sigma_0^2}\right), \tag{3.10}$$

and hence:

$$\ln p_0(z_2) = -\frac{(z_{22} - f(z_{21}))^2}{2} - \frac{z_{21}^2}{2\sigma_0^2} + \ln\left(\frac{1}{2\pi\sigma_0}\right). \tag{3.11}$$

For the first order term, we have:

$$\frac{\partial}{\partial \epsilon} \ln p_\epsilon(z_2)\Big|_{\epsilon=0} = \frac{1}{p_0(z_2)} \frac{\partial}{\partial \epsilon} p_\epsilon(z_2)\Big|_{\epsilon=0}. \tag{3.12}$$

This seemingly simple example, which has come to be called Witsenhausen's counter-example, turned out to be extremely hard. It is still outstanding after 30 years. However, new emerging applications and the necessity of looking back at some fundamental obstacles in designing decentralized stochastic strategies have recently inspired some new research on this example. In [1], a neural network, trained by stochastic approximation techniques, was used in order to approximate the optimal strategies. Also it was recently asserted in [4] that a global optimum would be achieved by searching directly in the strategy space using the generalized step functions to approximate the strategies.

In Witsenhausen's problem, the non-classical nature of the information pattern is a result of the fact that the information available to the first station is completely inaccessible for the second station. However, recent advances in computing and communication technologies make it possible for the stations in many decentralized systems to communicate different pieces of information. But the communications can never be perfect and there is always some uncertainty involved. Unfortunately, such uncertainty will again induce a non-classical nature on the information pattern of the system.

In [5], Witsenhausen's problem was reformulated in such a way that the first station could communicate its information with the second station through a noisy channel. It was shown, that as long as there is noise in transmission, the main difficulties will persist. Specifically, the cost might still be non-convex with respect to the strategies. However, when the transmission noise intensities are small, we would expect the optimal strategies to be very close to the corresponding strategies for the noiseless transmission case.

In the next section, we formulate the problem and discuss some of the results obtained in [5]. In Section 3, we approximate the cost by expanding it in terms of the small transmission noise intensity. In Section 4, we use a variational approach in order to find a necessary condition for the strategies which minimize the approximated cost. As we shall see, we will actually have a singular optimization problem. We will then show that asymptotically optimal strategies may still be linear with slightly different coefficients than the corresponding strategies for the noiseless transmission case. Finally, in the last section, we will have our concluding remarks.

2 Problem Description

Consider a two-stage stochastic problem with the following state equations:

$$x_1 = x_0 + u_1 \tag{2.1}$$
$$x_2 = x_1 - u_2, \tag{2.2}$$

where x_0 is the initial state, which is assumed to be a zero mean Gaussian random variable with variance σ_0^2. The information pattern of the system is specified by the following output equations:

$$z_1 = x_0 \tag{2.3}$$
$$z_2 = \begin{bmatrix} x_0 + v_t \\ x_0 + u_1 + v_2 \end{bmatrix} \triangleq \begin{bmatrix} z_{21} \\ z_{22} \end{bmatrix}, \tag{2.4}$$

where v_2 is the measurement noise for the second station, which is also assumed to be a zero mean Gaussian random variable with unit variance. As we can see, the information available to the first station is being transmitted to the second station and the communication uncertainty is modeled by an additive Gaussian noise $v_t \sim \mathcal{N}(0, \epsilon^2)$. Also, x_0, v_2 and v_t are all assumed to be independent of each other.

It is clear that we have simply modeled the received information signal as the transmitted signal plus a Gaussian transmission noise. While this model is realistic for analog communication systems, it may not be well justified when digital communication is used. In digital communication systems, the signal is quantized, coded and sent through the channel. Still, the channel noise may realistically be assumed to be additive and Gaussian, but sophisticated modulation and coding schemes make it difficult to assume a simple additive Gaussian uncertainty for the received information signal. However, if we try to incorporate the quantization effects along with the error probability distribution for some *good* coding and modulation schemes in order to model the communication uncertainties, we will end up with models which could still be approximated, to some degree, by simple additive Gaussian models. On the other hand, since there are already major difficulties in dealing with decentralized non-classical information patterns, using more complex models for communication uncertainties may not seem very reasonable at this point. Furthermore, we believe that the results obtained under such a simplifying assumption would still serve as a guideline for finding the true nature of decentralized strategies.

The objective is now to design:

$$u_1 = \gamma_1(z_1) \tag{2.5}$$
$$u_2 = \gamma_2(z_2), \tag{2.6}$$

in order to minimize the following cost function:

$$J = E\left[k^2 u_1^2 + x_2^2\right], \tag{2.7}$$

where $k^2 > 0$ is a given constant. We see that the first controller has perfect information but its action is costly. In contrast, the second controller has inexpensive control but noisy information. Since the second station does not know what the first station knew, due to the transmission noise, we don't have perfect recall and hence we still have a non-classical pattern. If there was no transmission noise, we would have a classical information pattern for which the unique optimal strategies are known to be linear in the information.

For simplicity, lets define:

$$f(z_1) \triangleq z_1 + \gamma_1(z_1) = x_0 + u_1 \tag{2.8}$$

On the other hand:

$$\frac{\partial}{\partial\epsilon}p_\epsilon(z_2)\bigg|_{\epsilon=0} =$$

$$\int_{-\infty}^{+\infty}\frac{\partial}{\partial\epsilon}\left\{\frac{1}{\sqrt{2\pi}}e^{-\frac{(z_{22}-f_\epsilon(y))^2}{2}}\frac{1}{\sqrt{2\pi}\sigma_0}e^{-\frac{(z_{21}+\epsilon y)^2}{2\sigma_0^2}}\right\}\bigg|_{\epsilon=0}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy$$

$$=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}(z_{22}-f(z_{21}))\,yf'(z_{21})\,e^{-\frac{(z_{22}-f(z_{21}))^2}{2}}$$

$$\frac{1}{\sqrt{2\pi}\sigma_0}e^{-\frac{z_{21}^2}{2\sigma_0^2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy+\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(z_{22}-f(z_{21}))^2}{2}}$$

$$\frac{1}{\sqrt{2\pi}\sigma_0}\left(-\frac{z_{21}}{\sigma_0^2}\right)ye^{-\frac{z_{21}^2}{2\sigma_0^2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy=0. \quad (3.13)$$

Therefore:

$$\frac{\partial}{\partial\epsilon}\ln p_\epsilon(z_2)\bigg|_{\epsilon=0}=0. \quad (3.14)$$

This result is not unexpected, because we would expect the behavior of $p_\epsilon(z_2)$ only to depend on the variance of the Gaussian transmission noise, i.e., ϵ^2. Using (3.14), we can now obtain the second order term as:

$$\frac{\partial^2}{\partial\epsilon^2}\ln p_\epsilon(z_2)\bigg|_{\epsilon=0}=\frac{1}{p_0(z_2)}\frac{\partial^2}{\partial\epsilon^2}p_\epsilon(z_2)\bigg|_{\epsilon=0}. \quad (3.15)$$

After some tedious but straightforward manipulations, we will get:

$$\frac{\partial^2}{\partial\epsilon^2}\ln p_\epsilon(z_2)\bigg|_{\epsilon=0}=$$

$$-f'^2(z_{21})+f''(z_{21})(z_{22}-f(z_{21}))+f'^2(z_{21})(z_{22}-f(z_{21}))^2$$

$$+2f'(z_{21})(z_{22}-f(z_{21}))\left(-\frac{z_{21}}{\sigma_0^2}\right)-\frac{1}{2\sigma_0^2}+\frac{z_{21}^2}{\sigma_0^4}. \quad (3.16)$$

We can now obtain a second order approximation for $\ln p_\epsilon(z_2)$ by substituting the corresponding terms from (3.11), (3.14) and (3.16) back into the expansion (3.6). In the next step, we substitute the expansion for $\ln p_\epsilon(z_2)$ in (2.12) in order to find the corresponding expansion for $g^*(z_2)$. Remember that $g^*(z_2)$ is the optimal strategy for the second station assuming that the first station has a fixed strategy $\gamma_1(z_1)=f(z_1)-z_1$. We have:

$$g^*(z_2) = z_{22}+\frac{\partial}{\partial z_{22}}\ln p(z_2)\simeq z_{22}+\frac{\partial}{\partial z_{22}}\ln p_0(z_2)$$

$$+\epsilon^2\frac{\partial}{\partial z_{22}}\left(\frac{\partial^2}{\partial\epsilon^2}\ln p_\epsilon(z_2)\bigg|_{\epsilon=0}\right)=z_{22}-(z_{22}-f(z_{21}))+$$

$$\epsilon^2\left[f''(z_{21})+2f'^2(z_{21})(z_{22}-f(z_{21}))+2f'(z_{21})\left(-\frac{z_{21}}{\sigma_0^2}\right)\right]. \quad (3.17)$$

Our goal is to get an expansion for the cost, which is in the form (2.13). Using the expansion for $g^*(z_2)$ from (3.17), we will have:

$$E\left[(g^*(z_2))^2\right]\simeq E\left[(f(z_{21}))^2\right]+2\epsilon^2E\left[f(z_{21})(f''(z_{21})\right.$$

$$\left.+2f'^2(z_{21})(z_{22}-f(z_{21}))+2f'(z_{21})(-z_{21}/\sigma_0^2))\right], \quad (3.18)$$

where we have neglected the fourth order term in ϵ. Substituting this expansion back in (2.13), we will obtain the following expansion for the cost:

$$J^*(f) = k^2E\left[(z_1-f(z_1))^2\right]+E\left[(f(z_1))^2\right]$$

$$-E\left[(f(z_{21}))^2\right]-2\epsilon^2E\left[f(z_{21})(f''(z_{21})\right.$$

$$\left.+2f'^2(z_{21})(z_{22}-f(z_{21}))+2f'(z_{21})(-z_{21}/\sigma_0^2))\right]. \quad (3.19)$$

Note that when the transmission is noiseless, i.e., $\epsilon=0$ and therefore $z_{21}=z_1$, we have:

$$J^*(f)=k^2E\left[(z_1-f(z_1))^2\right], \quad (3.20)$$

and $f(z_1)=z_1$ is the obvious unique optimal solution.

The above expansion, however, is not exactly in our desired form yet. This is because the third term on the right hand side, which is an average over z_{21}, still depends on ϵ. We shall now rewrite the expansion in (3.19) by explicitly expressing the expectations based on the corresponding probability densities:

$$J^*(f)=\int_{-\infty}^{+\infty}\left[k^2(t-f(t))^2+f^2(t)\right]\frac{1}{\sqrt{2\pi}\sigma_0}e^{-\frac{t^2}{2\sigma_0^2}}dt$$

$$-\int_{-\infty}^{+\infty}\left[f^2(t)+2\epsilon^2\left(f(t)f''(t)-2f(t)f'(t)\frac{t}{\sigma_0^2}\right)\right]$$

$$\frac{1}{\sqrt{2\pi(\sigma_0^2+\epsilon^2)}}e^{-\frac{t^2}{2(\sigma_0^2+\epsilon^2)}}dt-\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}4\epsilon^2f(t)f'^2(t)$$

$$(\tau-f(t))\frac{1}{\sqrt{2\pi}}e^{-\frac{(\tau-f(t))^2}{2}}\frac{1}{\sqrt{2\pi}\sigma_0}e^{-\frac{t^2}{2\sigma_0^2}}dtd\tau, \quad (3.21)$$

where we have substituted $p(z_2)=p(z_{22},z_{21})\simeq p_0(z_2)$ in the third term, since the higher order terms would be multiplied by ϵ^2 and then would be neglected. Now, the third term turns out to be zero, because:

$$\int_{-\infty}^{+\infty}4\epsilon^2f(t)f'^2(t)\frac{1}{\sqrt{2\pi}\sigma_0}e^{-\frac{t^2}{2\sigma_0^2}}.$$

$$\left(\int_{-\infty}^{+\infty}(\tau-f(t))\frac{1}{\sqrt{2\pi}}e^{-\frac{(\tau-f(t))^2}{2}}d\tau\right)dt=0. \quad (3.22)$$

On the other hand, we can expand the probability density of z_{21} up to the second order in ϵ:

$$\frac{1}{\sqrt{2\pi(\sigma_0^2+\epsilon^2)}}e^{-\frac{t^2}{2(\sigma_0^2+\epsilon^2)}}\simeq$$

$$\frac{1}{\sqrt{2\pi}\sigma_0}e^{-\frac{t^2}{2\sigma_0^2}}+\epsilon^2\frac{1}{\sqrt{2\pi}\sigma_0^5}(t^2-\sigma_0^2)e^{-\frac{t^2}{2\sigma_0^2}} \quad (3.23)$$

Substituting (3.22) and the above expansion back in (3.21) and neglecting the higher order terms in ϵ, we can finally get the following expansion for the cost:

$$J^*(f) = \int_{-\infty}^{+\infty}\left[k^2(t-f(t))^2\right]\frac{1}{\sqrt{2\pi}\sigma_0}e^{-\frac{t^2}{2\sigma_0^2}}dt$$

$$+\epsilon^2\int_{-\infty}^{+\infty}\left[4f(t)f'(t)\frac{t}{\sigma_0^2}-2f(t)f''(t)\right.$$
$$\left.+f^2(t)\frac{\sigma_0^2-t^2}{\sigma_0^4}\right]\frac{1}{\sqrt{2\pi}\sigma_0}e^{-\frac{t^2}{2\sigma_0^2}}dt \triangleq J_0^* + \epsilon^2 J_1^*. \quad (3.24)$$

The objective is now to obtain the function f which minimizes the above approximated cost. In the next section, we will use a variational approach in order to find the necessary conditions for such a function.

4 Minimizing the Approximated Cost

So far, we have obtained an expansion for the cost assuming that the transmission noise intensity is small. We have approximated the cost by including only up to the second order term in ϵ. We should now try to minimize this approximated cost and find the optimal f^*. Obviously, the corresponding optimal strategy would be valid only for a small transmission noise intensity. However, it would still be very helpful for the analysis of the behavior of the optimal strategies when we deviate a little bit from the classical information pattern by introducing a small communication uncertainty.

We now use the Hamiltonian approach in order to find the necessary conditions for the function $f(t)$, which minimizes our approximated cost. For simplicity, let's denote:

$$x_1(t) \triangleq f(t), \qquad x_2(t) \triangleq \dot{x}_1(t) = f'(t)$$
$$u(t) \triangleq \dot{x}_2(t) = \ddot{x}_1(t) = f''(t), \qquad p(t) \triangleq \frac{1}{\sqrt{2\pi}\sigma_0}e^{-\frac{t^2}{2\sigma_0^2}}$$

The Hamiltonian is then defined as [2]:

$$\mathcal{H}=k^2(t-x_1(t))^2p(t)+\epsilon^2\left(4x_1(t)x_2(t)\frac{t}{\sigma_0^2}-2x_1(t)u(t)\right.$$
$$\left.+x_1^2(t)\frac{\sigma_0^2-t^2}{\sigma_0^4}\right)p(t)+\lambda_1(t)x_2(t)+\lambda_2(t)u(t), \qquad (4.1)$$

where λ_1 and λ_2 are the Lagrange multipliers which should satisfy:

$$\dot{\lambda}_1(t)=-\mathcal{H}_{x_1}=\left(2k^2(t-x_1(t))-4\epsilon^2x_2(t)\frac{t}{\sigma_0^2}\right.$$
$$\left.-2\epsilon^2x_1(t)\frac{\sigma_0^2-t^2}{\sigma_0^4}+2\epsilon^2u(t)\right)p(t) \quad (4.2)$$
$$\dot{\lambda}_2(t)=-\mathcal{H}_{x_2}=-4\epsilon^2x_1(t)\frac{t}{\sigma_0^2}p(t)-\lambda_1(t). \qquad (4.3)$$

But as we can see, the Hamiltonian is linear in $u(t)$ and we actually have a *singular* optimization problem. The singular surface will be characterized by setting $\mathcal{H}_u$ and its derivatives with respect to t equal to zero, that is:

$$\mathcal{H}_u=-2\epsilon^2x_1(t)p(t)+\lambda_2(t)=0, \qquad (4.4)$$

and:

$$\frac{d}{dt}\mathcal{H}_u=-2\epsilon^2\dot{x}_1(t)p(t)-2\epsilon^2x_1(t)\dot{p}(t)+\dot{\lambda}_2(t)=0. \quad (4.5)$$

Substituting $\dot{p}(t) = -\frac{t}{\sigma_0^2}p(t)$ and also $\dot{\lambda}_2$ from (4.3), we will get:

$$\frac{d}{dt}\mathcal{H}_u=-2\epsilon^2x_2(t)p(t)-2\epsilon^2x_1(t)\frac{t}{\sigma_0^2}p(t)-\lambda_1(t)=0. \qquad (4.6)$$

Differentiating again and substituting $\dot{\lambda}_1$ from (4.2), we will have:

$$\frac{d^2}{dt^2}\mathcal{H}_u=\left(-4\epsilon^2u(t)+4\epsilon^2\frac{t}{\sigma_0^2}x_2(t)-2k^2(t-x_1(t))\right)p(t)=0. \qquad (4.7)$$

Therefore, the corresponding $u(t)$ on the singular surface is:

$$u(t)=x_2(t)\frac{t}{\sigma_0^2}-\frac{k^2}{2\epsilon^2}(t-x_1(t)). \qquad (4.8)$$

Note that the first order generalized Legendre-Clebsch condition, which is a necessary condition for $u(t)$ to be minimizing on the singular surface, is also satisfied, namely:

$$\frac{\partial}{\partial u}\left(\frac{d^2}{dt^2}\mathcal{H}_u\right)\leq 0, \qquad (4.9)$$

Therefore, the corresponding $x_1(t)$ and $x_2(t)$, which minimize our approximated cost, should necessarily satisfy the following differential equations:

$$\dot{x}_1(t) = x_2(t) \qquad (4.10)$$
$$\dot{x}_2(t) = x_2(t)\frac{t}{\sigma_0^2}-\frac{k^2}{2\epsilon^2}(t-x_1(t)) \qquad (4.11)$$

Since ϵ is assumed to be small, we may assume the following form in order to obtain the solutions for the above differential equations:

$$x_1(t) = a_0(t)+\epsilon^2a_2(t)+\epsilon^4a_4(t)+\ldots \quad (4.12)$$
$$x_2(t) = b_0(t)+\epsilon^2b_2(t)+\epsilon^4b_4(t)+\ldots \quad (4.13)$$

Interestingly enough, by substituting the above x_1 and x_2 back into the differential equations and comparing the coefficients of the terms with the same order in ϵ, we will get:

$$x_1(t)=\left[1-\frac{2\epsilon^2}{k^2\sigma_0^2}+\left(\frac{2\epsilon^2}{k^2\sigma_0^2}\right)^2-\left(\frac{2\epsilon^2}{k^2\sigma_0^2}\right)^3+\ldots\right]t=\frac{t}{\left(1+\frac{2\epsilon^2}{k^2\sigma_0^2}\right)}. \quad (4.14)$$

Back to our original notation, we indeed have:

$$f(z_1)=\frac{z_1}{\left(1+\frac{2\epsilon^2}{k^2\sigma_0^2}\right)}. \qquad (4.15)$$

As we can see, the solution is still linear with a coefficient which is slightly different than the corresponding coefficient for the noiseless transmission case. Remember that $f(z_1) = z_1$ is the optimal solution when there is no transmission noise and note that for $\epsilon = 0$ in (4.15), we get exactly the same solution, as we would expect. Also note that as the value of $k^2\sigma_0^2$ increases, the above solution approaches $f(z_1) = z_1$. In other words, increasing $k^2\sigma_0^2$ has an effect similar to decreasing the communication uncertainty. Given the above function $f(z_1)$, the corresponding $g^*(z_2)$ can easily be obtained using (2.11). Note that it will also be linear because of the Gaussian assumption for the underlying uncertainties.

In fact, we would expect the optimal strategies to be linear. As we mentioned in Section 1, linear strategies were shown to be asymptotically optimal for the Witsenhausen example when the uncertainty of the information available to the first station is small [3]. In this paper, however, we have considered a reformulation of Witsenhausen's problem where the first station sends its information to the second station through a low noise channel. These two scenario are somewhat similar. Namely, in both scenarios, the second station can determine the information available to the first station fairly accurately. Specifically, in the first scenario, the second station almost knows z_1 because of its small uncertainty, while in the second scenario, it can determine z_1 from the information that is transmitted through a low noise channel.

We would also expect the optimal strategies to approach the corresponding strategies for the noiseless transmission case as the value of z_1 and, in some sense, the *signal to noise ratio* increases. This doesn't seem to happen in the solution (4.15). One may justify this by looking at the exponential function in the cost. This function drives the integrand of the cost to zero exponentially fast for large z_1. Therefore, the structure of the cost does not force the optimal solution to approach $f(z_1) = z_1$ as z_1 increases.

Substituting $f(t)$ from (4.15) back into the cost (3.24), we obtain the corresponding value of the cost:

$$J^*(f) = \frac{1}{\left(1 + \frac{2\epsilon^2}{k^2\sigma_0^2}\right)^2}\left(2\epsilon^2 + \frac{4\epsilon^4}{k^2\sigma_0^2}\right) \simeq 2\epsilon^2 - \frac{4\epsilon^4}{k^2\sigma_0^2}, \tag{4.16}$$

The optimal cost for the noiseless transmission case is zero. But if we use $f(z_1) = z_1$ when the transmission is noisy, we get the following cost:

$$J^*(f) = 2\epsilon^2. \tag{4.17}$$

In other words, if we fix the strategies to be the optimal strategies for the noiseless transmission case while we introduce a small transmission noise, the increase in the cost will be proportional to the transmission noise intensity. However, if we use (4.15), we can indeed improve the cost by the fourth order in ϵ.

5 Concluding Remarks

We analyzed an example of a decentralized stochastic system. This example was a reformulation of the Witsenhausen counter-example where the first station was allowed to send its information to the second station through a noisy channel. The dynamics were linear, all the underlying uncertainties were assumed to be Gaussian and the cost was quadratic. However, the presence of the communication uncertainty had generated a non-classical information pattern. Therefore, in general, we would have a non-convex functional optimization problem.

We considered the case where the communication uncertainty was small. We followed an asymptotic approach where we approximated the cost based on its expansion in terms of the small transmission noise intensity. We showed how minimizing the approximated cost can be seen as a singular optimization problem. We then used a variational approach in order to find the necessary conditions for the asymptotically optimal strategies and showed that some reasonable linear strategies would actually satisfy those conditions. We also provided some intuitive explanations for the behavior of those linear strategies and obtained their corresponding cost.

All the derivations and the results in this paper show some of the difficulties involved in dealing with decentralized systems as soon as we deviate a little bit from a classical, or at least a partially nested, information pattern. On the other hand, even though we have modeled the communication uncertainty in the simplest possible way, we have tried to emphasize the role of communication uncertainties in generating such information patterns that are very difficult to handle.

Finally, even though the optimization problem is generally difficult for this class of systems, in some applications we might be able to exploit the specific structure of the system in order to obtain some reasonably good sup-optimal strategies, which would yield an acceptable performance.

References

[1] Baglietto, M. and Parisini, T.b and Zoppoli, R., "Nonlinear approximations for the solution of team optimal control problems", *Proceedings of CDC 1997*, Vol. 5, pp. 4592-4594, 1997.

[2] Bryson,Jr., A.E. and Ho, Y.C., *Applied optimal control*, Bristol,PA :Taylor & Francis, 1975.

[3] Castanon, D. A. and Sandell,Jr., N. R., "Signaling and uncertainty: a case study", *Proceedings of CDC 1978*, pp. 1140-1144, 1978.

[4] Lee, J. T. and Lau, E. and Ho, Y.C., "On the global optimum of the Witsenhausen counter-example", *Draft version*, 1998.

[5] Shoarinejad, K., Speyer, J. L. and Kanellakopoulos, I., "Decentralized control with noisy transmission of information", *Proceedings of ACC 1999*, pp. 1441-1444, 1999.

[6] Witsenhausen, H.S., "A counter-example in stochastic optimal control", *SIAM Journal of Control*, Vol. 6, pp. 131-147, 1968.

APPENDIX C

Two-Station Decentralized LQG Problem with Noisy Communication

K. Shoarinejad, J.D. Wolfe, J.L. Speyer, I. Kanellakopoulos

Proceedings of the 38th
Conference on Decision & Control
Phoenix, Arizona USA • December 1999

A Two-Station Decentralized LQG Problem with Noisy Communication*

Kambiz Shoarinejad
Ioannis Kanellakopoulos
UCLA Electrical Engineering
Box 951594, Los Angeles, CA 90095-1594
Tel: (310) 206-1022, Fax: (310) 206-1022
kambiz@ee.ucla.edu, ioannis@ee.ucla.edu

Jonathan D. Wolfe
Jason L. Speyer
UCLA Mechanical and Aerospace Engineering
Box 951597, Los Angeles, CA 90095-1597
Tel: (310) 206-4451, Fax: (310) 206-2302
wolfe@seas.ucla.edu, speyer@seas.ucla.edu

Abstract

We consider a two-station decentralized Linear Quadratic Gaussian problem, where the stations are allowed to communicate some pieces of information. We investigate a possible sub-optimal approach where the controls are obtained based on two separate centralized problems. Various cases will be considered in which the two stations communicate their measurements, their controls or estimates, and their estimation residuals through noisy channels. We will mainly focus on the closed-loop stability properties. We will show that even if the stations communicate all their measurements through low noise or even noiseless channels, the controls obtained from the two centralized LQG problems may fail to stabilize the closed-loop decentralized system.

1 Introduction

One of the most challenging problems for control engineers is to design controllers for large scale decentralized systems which are composed of a large number of spatially distributed interconnected subsystems. Uninhabited Air Vehicles (UAV's) flying in formation and Automated Vehicles driving in platoons are two examples. Also, recent advances in computing and communication technologies have introduced many new applications where dynamic systems would form spatial networks. However, there are still major difficulties in designing controllers for such systems that could achieve some specified level of performance.

It is always difficult to find decentralized stabilizing controllers as soon as the system has unstable fixed modes [3]. Incorporating uncertainties makes the problem even more difficult. This can be seen in a seemingly simple counterexample introduced by Witsenhausen in 1968, whose solution remains an open problem today. Witsenhausen showed that finding the optimal decentralized strategies, even for a very simple two-stage problem with linear dynamics, Gaussian uncertainties and quadratic cost, could be extremely hard as soon as the information pattern becomes non-classical.

In defining a decentralized linear quadratic Gaussian problem, we will assume all stations have linear dynamics and all uncertainties are modeled as Gaussian processes. Moreover, each local controller only has access to its own local information, which includes its own measurements and possibly information received through communication with other stations. Such a decentralized nature of information generally induces a non-classical information pattern for this class of problems. Therefore, except for some special structures, where the information pattern is actually a classical pattern [2], the optimal strategies are usually unknown. Some sub-optimal approaches , however, might be proposed. One such approach is to treat the problem as a collection of separate centralized problems. A motivation for this approach would become clearer if we assume that each station is allowed to communicate all its measurements through low noise communication channels with all the other stations. Even though a huge burden of computation and communication resources may be needed in this scenario, we would expect the controllers to be very close to the optimal stabilizing decentralized controllers.

In the next section, we formulate a simple two-station decentralized LQG problem. In Section 3, we discuss the above mentioned sub-optimal approach, where we propose a solution based on two separate centralized problems. In Section 4, we investigate the stability properties of our controllers in various scenarios. Namely, we first consider the case where the stations do not communicate at all. Then, we assume that the stations can communicate their state estimates or equivalently their controls. In these scenarios, as we shall see, there is little justification for our approach. But later, we will discuss the case where the stations are allowed to communicate their measurements. As we mentioned, our approach seems very reasonable for this scenario, at least when the transmission noise intensities are assumed to be small. However, as one of the main results in this paper,

*This research was supported in part by the National Science Foundation under Grant ECS-9502945, Air Force Office of Scientific Research under Grant F49620-97-1-0272 and Office of Naval Research under Award N00014-97-1-0939

we will show that even in this case, our controllers may fail to stabilize the closed-loop system. This clearly contradicts our expectation. In another scenario, the stations will be allowed to communicate both their measurements and their controls. We will see how sending the controls will help us achieve at least closed-loop stability with our controllers. Then, we assume that the stations communicate only their estimation residuals. We will show that when the transmission noise intensities are small, sending estimation residuals would be enough to achieve closed-loop stability in our sub-optimal approach. Our concluding remarks appear in the final section.

2 Problem Statement

Consider the following decentralized linear system with two stations:

$$\dot{x}(t) = Ax(t) + B_1u^1(t) + B_2u^2(t) + w(t) \quad (2.1)$$
$$z^1(t) = H^1x(t) + v^1(t) \quad (2.2)$$
$$z^2(t) = H^2x(t) + v^2(t), \quad (2.3)$$

where $x(t) \in \mathcal{R}^n$ is the global state vector, $u^1(t) \in \mathcal{R}^{m_1}$ and $z^1(t) \in \mathcal{R}^{r_1}$ are the control and the information vectors for the first station and $u^2(t) \in \mathcal{R}^{m_2}$ and $z^2(t) \in \mathcal{R}^{r_2}$ are the control and the information vectors for the second station. The process noise and the information noises are denoted by $w(t)$, $v^1(t)$ and $v^2(t)$ respectively, which are all assumed to be zero mean white Gaussian with intensity matrices W, V^1 and V^2. They are also assumed to be mutually independent and independent of the initial state. Note that we distinguish between measurement and information, simply because of the fact that the information vector for a station may also include the transmitted measurements of the other station.

The original objective is to find $u^1 = u^1(z^1)$ and $u^2 = u^2(z^2)$ in order to minimize the following cost:

$$J = \lim_{T\to\infty} \frac{1}{T} E\left[\int_0^T \left(x^T Q x + {u^1}^T R_1 u^1 + {u^2}^T R_2 u^2\right) dt\right]. \quad (2.4)$$

Since the stations in general have access to different information, we have a non-classical information pattern. Moreover, the information pattern is not partially nested. That is, the information available to each station is being affected by the control action of the other station, while there is no way for that station to obtain any information about those control actions. Therefore, in general, we will have a non-convex functional optimization problem, the solutions of which are usually very difficult to find.

One possible sub-optimal approach is to solve two separate centralized problems. We will discuss this approach in the following sections. But there are two points that we need to mention now. As we shall see, in many cases, we are fixing the structure of our controllers only based on the centralized results. Even though this comes naturally out of our lack of knowledge about the structure of the decentralized controllers, it may well be justified for the case where the stations communicate all their measurements through low noise channels. The other point is our choice of model for the uncertainty in the transmitted information. We simply model the received information signal as the transmitted signal plus a Gaussian transmission noise. While this model is realistic for analog communication systems, it may not be well justified when digital communication is used. Namely, in digital communication systems, the signal is quantized, coded and sent through the channel. The channel noise may still be assumed to be additive and Gaussian, but sophisticated modulation and coding schemes make it difficult to assume a simple additive Gaussian uncertainty for the received information signal. However, if we try to incorporate the quantization effects along with the error probability distribution for some *good* coding and modulation schemes in order to model the communication uncertainties, we will end up with models which could still be approximated, to some degree, by simple additive Gaussian models. On the other hand, since there are already major difficulties in dealing with decentralized non-classical information patterns, using more complex models for communication uncertainties does not seem very reasonable at this point. Furthermore, we believe the results obtained under such a simplifying assumption would still be helpful in giving us insight towards the true nature of decentralized controllers.

3 A Sub-optimal Approach

One possible sub-optimal approach in dealing with decentralized problems is to decompose them into several centralized problems in a reasonable fashion. One of our main objectives in this paper is to investigate such an approach and elaborate more on some of the important properties of the controllers under various communication scenarios among the stations.

Consider the system (2.1) again. We would like to design the controls based on two centralized LQG problems. Namely, let each station pretend that it has access to both of the controls while it only has access to its own information. In other words, the i-th station ($i = 1, 2$) wants to design $u_1^i = u_1^i(z^i)$ and $u_2^i = u_2^i(z^i)$ in order to minimize the following cost:

$$J_i = \lim_{T\to\infty} \frac{1}{T} E\left[\int_0^T \left(x^T Q x + {u_1^i}^T R_1 u_1^i + {u_2^i}^T R_2 u_2^i\right) dt\right]. \quad (3.1)$$

From the well-known centralized LQG results [1], the optimal controls can be obtained as:

$$\begin{bmatrix} u_1^i(t) \\ u_2^i(t) \end{bmatrix} = \begin{bmatrix} -R_1^{-1}B_1^T\Pi\hat{x}^i(t) \\ -R_2^{-1}B_2^T\Pi\hat{x}^i(t) \end{bmatrix} = \begin{bmatrix} -K_1\hat{x}^i(t) \\ -K_2\hat{x}^i(t) \end{bmatrix}, \quad i = 1, 2, \quad (3.2)$$

where Π is obtained from the steady-state control Riccati equation:

$$-\Pi A - A^T\Pi + \Pi\left(B_1R_1^{-1}B_1^T + B_2R_2^{-1}B_2^T\right)\Pi - Q = 0\,, \tag{3.3}$$

and $\hat{x}^i$ is the local state estimate in the i-th station:

$$\dot{\hat{x}}^i(t) = A\hat{x}^i(t) + B_1u_1^i(t) + B_2u_2^i(t) + L_i\left(z^i(t) - H^i\hat{x}^i(t)\right), \tag{3.4}$$

The estimator gain is obtained as:

$$L_i \triangleq P_i(H^i)^T\left(V^i\right)^{-1}, \quad i = 1,2\,, \tag{3.5}$$

where P_i is the solution to the corresponding steady-state filter Riccati equation:

$$AP_i + P_iA^T - P_i(H^i)^T\left(V^i\right)^{-1}H^iP_i + W = 0, \quad i = 1,2\,. \tag{3.6}$$

Note that the only difference in the two centralized problems comes from the fact that the stations have access to different information, i.e. from the matrix H^i and the noise intensity matrix V^i.

After solving the two centralized problems, u_1^1 and u_2^2 will be applied to the decentralized system. Obviously, there is no reason for these controllers to be optimal for the decentralized system. Also they are not guaranteed to preserve any level of performance including even the closed-loop stability. However, in some cases, where the stations are allowed to communicate some pieces of information through low noise channels, we would expect the local stations to generate very similar controllers, which in turn are expected to be very close to the decentralized optimal controllers.

4 Closed-Loop Stability

Achieving closed-loop stability is one of the most important performance properties that we would desire for our controllers. On the other hand, the centralized LQG controllers will always stabilize the system under some *detectability* and *stabilizability* conditions. But in general, there is no reason to guarantee closed-loop stability if we apply the same centralized controls to the decentralized system. In this section, we will investigate the closed-loop stability properties of our controllers in various situations, where the stations communicate different pieces of information. Note that in some cases, based on the available information for each station, we may modify the estimators and hence deviate a little bit from the original centralized LQG solutions. In such cases, we will instead be looking at general *linear estimate linear feedback* structures.

In order to analyze the dynamics of the closed loop system, we define the local estimation errors and the difference between the local estimates respectively as:

$$e_1(t) \triangleq x(t) - \hat{x}^1(t) \tag{4.1}$$

$$e_2(t) \triangleq x(t) - \hat{x}^2(t) \tag{4.2}$$

$$e_{12}(t) \triangleq \hat{x}^1(t) - \hat{x}^2(t). \tag{4.3}$$

It is straightforward to obtain:

$$\begin{aligned}\dot{x} = Ax - B_1K_1\hat{x}^1 - B_2K_2\hat{x}^2 + w = (A - B_1K_1 - B_2K_2)\,x\\ + (B_1K_1 + B_2K_2)\,e_2 - B_1K_1e_{12} + w\end{aligned} \tag{4.4}$$

$$\dot{e}_2 = \left(A - L_2H^2\right)e_2 - B_1K_1e_{12} + w - L_2v^2 \tag{4.5}$$

$$\begin{aligned}\dot{e}_{12} = \left(A - B_1K_1 - B_2K_2 - L_1H^1\right)e_{12}\\ + \left(L_1H^1 - L_2H^2\right)e_2 + L_1v^1 - L_2v^2,\end{aligned} \tag{4.6}$$

Hence, the closed-loop system dynamics can be written as follows:

$$\begin{aligned}\begin{bmatrix}\dot{x}\\ \dot{e}_2\\ \dot{e}_{12}\end{bmatrix} = \begin{bmatrix}A - B_1K_1 - B_2K_2 & B_1K_1 + B_2K_2 & -B_1K_1\\ 0 & A - L_2H^2 & -B_1K_1\\ 0 & L_1H^1 - L_2H^2 & A - B_1K_1 - B_2K_2 - L_1H^1\end{bmatrix}\begin{bmatrix}x\\ e_2\\ e_{12}\end{bmatrix} + \begin{bmatrix}I & 0 & 0\\ I & 0 & -L_2\\ 0 & L_1 & -L_2\end{bmatrix}\begin{bmatrix}w\\ v^1\\ v^2\end{bmatrix}.\end{aligned} \tag{4.7}$$

4.1 No Transmission

Assume that each station only has access to its own measurements, i.e., there is no communication between the stations. In this case, the closed-loop dynamics are in the form (4.7) where H^1 and H^2 are the corresponding measurement matrices for the stations, while v^1 and v^2 simply denote the measurement uncertainties.

Let's assume that the stations have the same measurement characteristics. Then it is clear from (4.7) that in order to have a stable closed-loop system, we need to have stable feedback dynamics along with stable local estimators and compensators. We conjecture that these stability properties are sufficient for the closed-loop stability even if the stations do not have identical measurements. But to achieve such stability properties, we need the global state to be detectable from each local station. This condition, however, is a very strong condition for a decentralized system. In most decentralized systems the global state can not be detectable from all individual stations. Moreover, even if such a strong condition is satisfied, we still do not have any good justification for our sub-optimal approach in this case. There is really no reason to expect the two centralized controllers to have a good performance if they are applied to the decentralized system.

4.2 Control (Estimate) Transmission

In this scenario, the stations communicate only their controls. In other words, each station has access to its own local measurements and the transmitted control of the other station. As we have already mentioned, the communication uncertainties are simply modeled as additive Gaussian noises. Also all the communications are assumed to be instantaneous. Therefore, the information available to the first station is:

$$z_1 = H_1x + v_1, \qquad u_2(t) + v_{t2}(t), \tag{4.8}$$

while the second station has access to the following information:

$$z_2 = H_2 x + v_2, \qquad u_1(t) + v_{t1}(t), \tag{4.9}$$

where v_{t1} and v_{t2} are the corresponding transmission noises. Each station now incorporates the received control of the other station in its local estimator. Namely, the local estimators are:

$$\dot{\hat{x}}^1 = A\hat{x}^1 + B_1 u_1 + B_2 u_2 + B_2 v_{t2} + L_1\left(z_1 - H_1 \hat{x}^1\right) \tag{4.10}$$

$$\dot{\hat{x}}^2 = A\hat{x}^2 + B_1 u_1 + B_1 v_{t1} + B_2 u_2 + L_2\left(z_2 - H_2 \hat{x}^2\right), \tag{4.11}$$

where:

$$L_1 \triangleq P_1 H_1^T V_1^{-1} \tag{4.12}$$

$$L_2 \triangleq P_2 H_2^T V_2^{-1} \tag{4.13}$$

and P_1 and P_2 are still the solutions to the corresponding Riccati equations. Note that P_1 and P_2 are not the local estimation error covariances anymore. The following controls are now applied to the decentralized system:

$$\begin{aligned} u_1(t) &= -R_1 B_1^T \Pi \hat{x}^1(t) = -K_1 \hat{x}^1(t) && (4.14)\\ u_2(t) &= -R_2 B_2^T \Pi \hat{x}^2(t) = -K_2 \hat{x}^2(t), && (4.15)\end{aligned}$$

where Π is the solution to the corresponding steady-state control Riccati equation. It is straightforward to obtain the dynamics of the closed-loop system:

$$\begin{bmatrix} \dot{x} \\ \dot{e}_1 \\ \dot{e}_2 \end{bmatrix} = \begin{bmatrix} A - B_1K_1 - B_2K_2 & B_1K_1 & B_2K_2 \\ 0 & A - L_1H_1 & 0 \\ 0 & 0 & A - L_2H_2 \end{bmatrix} \begin{bmatrix} x \\ e_1 \\ e_2 \end{bmatrix} + \begin{bmatrix} I & 0 & 0 \\ I & -L_1 & 0 \\ I & 0 & -L_2 \end{bmatrix} \begin{bmatrix} w \\ v^1 \\ v^2 \end{bmatrix} - B_2 \begin{bmatrix} 0 \\ v_{t2} \\ 0 \end{bmatrix} - B_1 \begin{bmatrix} 0 \\ 0 \\ v_{t1} \end{bmatrix}. \tag{4.16}$$

It is clear that the closed-loop system can be stabilized if the system is stabilizable using both stations and it is detectable from each individual station. As we mentioned earlier, this latter condition can not be satisfied in many decentralized systems. Also even if the control transmission is noiseless, there is still no reason to believe that these centralized controllers are, in any sense, close to the optimal decentralized controllers.

Note that communicating the local estimates is actually equivalent to communicating the controls. This is because we have a cooperative structure. That is, each station can be informed of the control strategy and specifically the estimator and feedback gains of the other station *a priori*. Therefore, the stations can simply calculate either the control or the estimate upon receiving the other.

Finally, note that we have incorporated the transmitted controls in the local estimators in a rather straightforward manner. Whether there are better ways to incorporate this new information is a problem to be addressed.

4.3 Measurement Transmission

Assume now that the stations can communicate all their measurements. In this case, the information available to the stations can be expressed as:

$$z^1 \triangleq \begin{bmatrix} z_1^1 \\ z_2^1 \end{bmatrix} = \begin{bmatrix} H_1 x + v_1 \\ H_2 x + v_2 + v_{21} \end{bmatrix} \triangleq Hx + v^1 \tag{4.17}$$

$$z^2 \triangleq \begin{bmatrix} z_1^2 \\ z_2^2 \end{bmatrix} = \begin{bmatrix} H_1 x + v_1 + v_{12} \\ H_2 x + v_2 \end{bmatrix} \triangleq Hx + v^2, \tag{4.18}$$

where $v_{12}(t)$ and $v_{21}(t)$ are independent transmission noises, which are also assumed to be independent of other underlying uncertainties in the system. Note that in this scenario, both stations have the same information matrix H. Therefore, there can not be any decentralized fixed mode in this case.

Similar to the previous cases, we solve two separate centralized LQG problems. For the first station we get:

$$\begin{bmatrix} u_1^1(t) \\ u_2^1(t) \end{bmatrix} = \begin{bmatrix} -R_1^{-1} B_1^T \Pi \hat{x}^1(t) \\ -R_2^{-1} B_2^T \Pi \hat{x}^1(t) \end{bmatrix} = \begin{bmatrix} -K_1 \hat{x}^1(t) \\ -K_2 \hat{x}^1(t) \end{bmatrix}, \tag{4.19}$$

where:

$$\dot{\hat{x}}^1 = A\hat{x}^1 + B_1 u_1^1 + B_2 u_2^1 + L_1\left(z^1 - H\hat{x}^1\right) \tag{4.20}$$

$$L_1 \triangleq P_1 H^T \left(V^1\right)^{-1} \tag{4.21}$$

$$AP_1 + P_1 A^T - P_1 H^T \left(V^1\right)^{-1} H P_1 + W = 0 \tag{4.22}$$

$$H \triangleq \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}, \quad V^1 \triangleq \begin{bmatrix} V_1 & 0 \\ 0 & V_2 + V_{21} \end{bmatrix}, \tag{4.23}$$

and for the second station:

$$\begin{bmatrix} u_1^2(t) \\ u_2^2(t) \end{bmatrix} = \begin{bmatrix} -R_1^{-1} B_1^T \Pi \hat{x}^2(t) \\ -R_2^{-1} B_2^T \Pi \hat{x}^2(t) \end{bmatrix} = \begin{bmatrix} -K_1 \hat{x}^2(t) \\ -K_2 \hat{x}^2(t) \end{bmatrix}, \tag{4.24}$$

where:

$$\dot{\hat{x}}^2 = A\hat{x}^2 + B_1 u_1^2 + B_2 u_2^2 + L_2\left(z^2 - H\hat{x}^2\right) \tag{4.25}$$

$$L_2 \triangleq P_2 H^T \left(V^2\right)^{-1} \tag{4.26}$$

$$AP_2 + P_2 A^T - P_2 H^T \left(V^2\right)^{-1} H P_2 + W = 0 \tag{4.27}$$

$$H \triangleq \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}, \quad V^2 \triangleq \begin{bmatrix} V_1 + V_{12} & 0 \\ 0 & V_2 \end{bmatrix}. \tag{4.28}$$

In this scenario, we have a very good justification for our sub-optimal approach. Namely, if the transmissions are noiseless, the two centralized problems would be identical. Therefore, we would expect our controllers to be the optimal decentralized controllers, which would preserve all the desired properties including the closed-loop stability. Furthermore, if the transmissions are noisy but the transmission noise intensities are small, we would still expect the controllers to be close to the optimal stabilizing decentralized controllers. In other words, we would not expect any drastic change in the behavior of the controlled decentralized system upon introducing some small transmission noise.

We shall now look at the closed-loop stability properties. It is easy to obtain the following closed-loop system dynamics which is valid for any linear estimate linear feedback structure:

$$\begin{bmatrix} \dot{x} \\ \dot{e}_2 \\ \dot{e}_{12} \end{bmatrix} = \begin{bmatrix} A-B_1K_1-B_2K_2 & B_1K_1+B_2K_2 & -B_1K_1 \\ 0 & A-L_2H & -B_1K_1 \\ 0 & (L_1-L_2)H & A-B_1K_1-B_2K_2-L_1H \end{bmatrix} \begin{bmatrix} x \\ e_2 \\ e_{12} \end{bmatrix} + \begin{bmatrix} I & 0 & 0 \\ I & 0 & -L_2 \\ 0 & L_1 & -L_2 \end{bmatrix} \begin{bmatrix} w \\ v^1 \\ v^2 \end{bmatrix}. \tag{4.29}$$

We notice that the closed-loop system matrix has an interesting structure. The first diagonal block matrix is simply the matrix associated with the feedback dynamics, which could be stabilized if the system is stabilizable using both control stations. The second diagonal block matrix could also be made stable under a simple detectability condition. That is, if the global state is detectable using both stations. Note that this is a much weaker condition than detectability from each individual station, which would be required if the stations did not communicate their measurements. The third diagonal block matrix, however, is the matrix corresponding to the *compensator* dynamics, which may not be stable.

This is a significant result. Let's assume that the transmission noise intensities are very small. Then the estimator gains would be almost the same and the closed-loop system matrix would be very close to a block upper-triangular matrix. We can see that if the compensator is unstable (which might be the case in many systems, especially those with a non-minimum phase structure), the closed-loop system will become unstable because of the unstable dynamics governing the difference between the estimates of the two local estimators. Actually, even when the transmissions are noiseless, there is still an unstable subsystem corresponding to e_{12}. This does not comply with our initial expectation. Note that there is no forcing input for this unstable subsystem, but any small nonzero e_{12} could propagate to infinity! Such a nonzero difference between the local estimates, which could be generated from any difference in the initial conditions of the local estimators, round off errors, etc., would again induce a non-classical information pattern.

4.4 Measurement and Control Transmission

We saw that if the stations communicate only their measurements, our specific sub-optimal controllers may not be able to stabilize the closed-loop system, even though they will yield the centralized optimal stabilizing controllers, in the limit, when the transmission noise intensities go to zero. In this section, we will see how transmitting the controls along with the measurements will help us stabilize the closed-loop system, using a similar sub-optimal approach.

As in the previous case, assume that the stations transmit their measurements through noisy channels, i.e.:

$$z^1(t) \triangleq \begin{bmatrix} z_1^1(t) \\ z_2^1(t) \end{bmatrix} = \begin{bmatrix} H_1x(t)+v_1(t) \\ H_2x(t)+v_2(t)+v_{21}(t) \end{bmatrix} \tag{4.30}$$

$$z^2(t) \triangleq \begin{bmatrix} z_1^2(t) \\ z_2^2(t) \end{bmatrix} = \begin{bmatrix} H_1x(t)+v_1(t)+v_{12}(t) \\ H_2x(t)+v_2(t) \end{bmatrix}. \tag{4.31}$$

Also assume that the stations communicate their controls. For a little more generality, lets assume that the communication uncertainties on the controls are modeled by an additive Gaussian uncertainty along with a scale-factor error. Namely, the first station also has access to $(I+\Delta_2)u^2(t) + v_{t2}(t)$, while the second station receives $(I+\Delta_1)u^1(t) + v_{t1}(t)$. Transmission noises $v_{t1}(t)$ and $v_{t2}(t)$ are assumed to be independent of each other and also independent of all other uncertainties in the system.

Similar to Section 4.2, each station incorporates the transmitted control of the other station in its local estimator. That is, the estimators are constructed in the following manner:

$$\dot{\hat{x}}^1 = A\hat{x}^1+B_1u^1+B_2(I+\Delta_2)u^2+B_2v_{t2}+L_1(z^1-H\hat{x}^1) \tag{4.32}$$

$$\dot{\hat{x}}^2 = A\hat{x}^2+B_1(I+\Delta_1)u^1+B_1v_{t1}+B_2u^2+L_2(z^2-H\hat{x}^2), \tag{4.33}$$

where:

$$L_1 \triangleq P_1H^T\left(V^1\right)^{-1} \tag{4.34}$$

$$L_2 \triangleq P_2H^T\left(V^2\right)^{-1} \tag{4.35}$$

and P_1 and P_2 are obtained from the same Riccati equations as before. Note that P_1 and P_2 are no longer the estimation error covariances. Using the same definitions for the error variables $e_1(t)$ and $e_2(t)$, the closed-loop dynamics may be written as:

$$\begin{bmatrix} \dot{x} \\ \dot{e}_1 \\ \dot{e}_2 \end{bmatrix} = \begin{bmatrix} A-B_1K_1-B_2K_2 & B_1K_1 & B_2K_2 \\ -B_2K_2\Delta_2 & A-L_1H & B_2K_2\Delta_2 \\ -B_1K_1\Delta_1 & B_1K_1\Delta_1 & A-L_2H \end{bmatrix} \begin{bmatrix} x \\ e_1 \\ e_2 \end{bmatrix} + \begin{bmatrix} I & 0 & 0 \\ I & -L_1 & 0 \\ I & 0 & -L_2 \end{bmatrix} \begin{bmatrix} w \\ v^1 \\ v^2 \end{bmatrix} - B_2 \begin{bmatrix} 0 \\ v_{t2} \\ 0 \end{bmatrix} - B_1 \begin{bmatrix} 0 \\ 0 \\ v_{t1} \end{bmatrix}. \tag{4.36}$$

As we can see, when the scale-factor errors Δ_1 and Δ_2 are small, the closed-loop system matrix is nearly block upper-triangular. The first diagonal block matrix can be made stable if the system is stabilizable using both stations. The second and the third diagonal block matrices can also be made stable if (A, H) is detectable.

We conclude that when the stations communicate their controls as well as their measurements, our sub-optimal approach will at least yield a stable closed-loop system, even if there is small scale-factor errors on the control transmissions.

4.5 Estimation Residuals Transmission

So far, we have seen that in order to design a set of sub-optimal stabilizing controllers by solving two centralized problems for a two-station decentralized system and under some reasonable stabilizability and detectability assumptions, the stations need to communicate both their measurements and controls.

In this section, we investigate the case where the stations communicate their estimation residuals instead of their measurements and controls. In other words, the first station has access to the following information:

$$z_1 = H_1 x + v_1, \quad \left(z_2 - H_2 \hat{x}^2\right) + v_{t2}, \tag{4.37}$$

while the information available to the second station is:

$$z_2 = H_2 x + v_2, \quad \left(z_1 - H_1 \hat{x}^1\right) + v_{t1}, \tag{4.38}$$

where v_{t1} and v_{t2} denote the transmission noises. In the previous cases, the linear structure of the estimators and the controllers naturally came out of the two centralized optimal control problems. In this case, however, we will impose a linear structure on our estimation and control such that each station will linearly incorporate the noisy residual of the other station, i.e., for the first station, we have:

$$u_1^1(t) = -K_1 \hat{x}^1(t), \quad u_2^1(t) = -K_2 \hat{x}^1(t) \tag{4.39}$$

$$\begin{aligned} \dot{\hat{x}}^1 = {} & A\hat{x}^1 + B_1 u_1^1 + B_2 u_2^1 + L_1^1 \left(z_1 - H_1 \hat{x}^1\right) \\ & + L_2^1 \left(z_2 - H_2 \hat{x}^2\right) + L_2^1 v_{t2}, \end{aligned} \tag{4.40}$$

while for the second station, we get:

$$u_2^1(t) = -K_1 \hat{x}^2(t), \quad u_2^2(t) = -K_2 \hat{x}^2(t) \tag{4.41}$$

$$\begin{aligned} \dot{\hat{x}}^2 = {} & A\hat{x}^2 + B_1 u_1^2 + B_2 u_2^2 + L_1^2 \left(z_1 - H_1 \hat{x}^1\right) \\ & + L_2^2 \left(z_2 - H_2 \hat{x}^2\right) + L_1^2 v_{t1}, \end{aligned} \tag{4.42}$$

The gains may now be obtained based on some optimality criteria. Note that when the transmission noises v_{t1} and v_{t2} are zero, the local estimators will have exactly the same structure. Therefore, we expect the estimators to have the same gains in the noiseless transmission case, regardless of how the gains are obtained. Also note that each station has linearly incorporated the received estimation residual of the other station. Even though this simplifies the problem, it is not necessarily the best way of incorporating this new piece of information.

Similarly to the previous cases, it is straightforward to obtain the closed-loop dynamics as the following:

$$\begin{aligned} \begin{bmatrix} \dot{x} \\ \dot{e}_2 \\ \dot{e}_{12} \end{bmatrix} = {} & \begin{bmatrix} A - B_1 K_1 - B_2 K_2 & B_1 K_1 + B_2 K_2 & -B_1 K_1 \\ 0 & A - L_1^2 H_1 - L_2^2 H_2 & L_1^2 H_1 - B_1 K_1 \\ 0 & \left(L_2^1 - L_2^2\right) H_2 & A - B_1 K_1 - B_2 K_2 - \left(L_1^1 - L_1^2\right) H_1 \end{bmatrix} \begin{bmatrix} x \\ e_2 \\ e_{12} \end{bmatrix} \\ & + \begin{bmatrix} I & 0 & 0 \\ I & -L_1^2 & -L_2^2 \\ 0 & \left(L_1^1 - L_1^2\right) & \left(L_2^1 - L_2^2\right) \end{bmatrix} \begin{bmatrix} w \\ v_1 \\ v_2 \end{bmatrix} - \begin{bmatrix} 0 \\ L_1^2 \\ L_1^2 \end{bmatrix} v_{t1} + \begin{bmatrix} 0 \\ 0 \\ L_2^1 \end{bmatrix} v_{t2}. \end{aligned} \tag{4.43}$$

As we can see, when the transmission noise intensities are small, the closed-loop system matrix will be close to a block upper-triangular matrix, which can easily be stabilized when the system is stabilizable using both stations and $\left(A, \left[\begin{array}{cc} H_1^T & H_2^T \end{array}\right]^T\right)$ is detectable. This shows us that in some sense, the estimation residuals are more valuable than the measurements and communicating the residuals is enough to stabilize the system by solving two centralized problems.

5 Concluding Remarks

A two station decentralized LQG problem was formulated, where the local controllers had to be designed based on some local information in order to minimize a single common cost. This problem generally has a non-classical information pattern and the optimal controls are usually unknown. One of the first possible sub-optimal approaches is to decompose the problem into separate centralized problems. In this paper, we investigated such an approach for different communication scenarios between the stations, namely, when the stations communicate their controls, their measurements or both, or their estimation residuals.

We showed that even though our approach is quite reasonable for the case where the stations communicate all their measurements, it may fail to stabilize the closed-loop system as soon as the compensator is unstable. Then, we showed how this difficulty can be removed if the stations either communicate both their measurements and their controls or communicate their estimation residuals. We should also mention that a similar problem can be formulated for discrete-time systems and similar results can be obtained.

All these results show some of the fundamental differences between the centralized and the decentralized structures. Moreover, we have tried to elaborate on the role of communication among the stations and the corresponding uncertainties. While many new applications for spatially distributed dynamic systems are emerging, there are still major difficulties that need to be addressed.

References

[1] Bryson Jr., A. E. and Ho, Y. C., *Applied optimal control*, Bristol,PA :Taylor & Francis, 1975.

[2] Speyer, J. L., "Computation and transmission requirements for a decentralized linear-quadratic-gaussian control problem", *IEEE Trans. on Automatic Control*, Vol. AC-24, pp. 266-9, 1979.

[3] Wang, S. H., Davison, E. J., "On the stabilization of decentralized control systems", *IEEE Trans. on Automatic Control*, Vol. AC-18, pp. 473-8, 1973.

APPENDIX D

A Generalized Least-Squares Fault Detection Filter

R.H. Chen and J.L. Speyer

INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. 2000; 14:747–757

A generalized least-squares fault detection filter

Robert H. Chen*,†, and Jason L. Speyer

Mechanical and Aerospace Engineering Department, University of California, Los Angeles, Los Angeles, CA 90095-1597, U.S.A.

SUMMARY

A fault detection and identification algorithm is determined from a generalization of the least-squares derivation of the Kalman filter. The objective of the filter is to monitor a single fault called the target fault and block other faults which are called nuisance faults. The filter is derived from solving a min–max problem with a generalized least-squares cost criterion which explicitly makes the residual sensitive to the target fault, but insensitive to the nuisance faults. It is shown that this filter approximates the properties of the classical fault detection filter such that in the limit where the weighting on the nuisance faults is zero, the generalized least-squares fault detection filter becomes equivalent to the unknown input observer where there exists a reduced-order filter. Filter designs can be obtained for both linear time-invariant and time-varying systems. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: fault detection and identification; unknown input observer; worst case design; time-varying system

1. INTRODUCTION

Any system under automatic control demands a high degree of system reliability. This requires a health monitoring system capable of detecting any plant, actuator and sensor fault as it occurs and identifying the faulty component. One approach, analytical redundancy which reduces the need for hardware redundancy, uses the modelled dynamic relationship between system inputs and measured system outputs to form a residual process used for detecting and identifying faults. A popular approach to analytical redundancy is the unknown input observer [1] which divides the faults into two groups: a single-target fault and possibly several nuisance faults. The nuisance faults are placed in an invariant subspace which is unobservable to the residual. Recently, approximate unknown input observers have been developed which have improved robustness to uncertainties and applicable to time-varying systems [2,3].

In this paper, a generalized least-squares fault detection filter, motivated by Chung and Speyer [2] and Bryson and Ho [4], is presented. A new least-squares problem with an indefinite cost

* Correspondence to: Robert H. Chen, Mechanical and Aerospace Engineering Department, University of California, Los Angeles, California 90095-1597, U.S.A.
† E-mail: chrobert@talus.seas.ucla.edu

Contract/grant sponsor: Air Force Office of Scientific Research; contract/grant number: F49620-97-1-0272
Contract/grant sponsor: California Department of Transportation; contract/grant number: 65A0013, MOU315

criterion is formulated as a min–max problem by generalizing the least-squares derivation of the Kalman filter [4] and allowing the explicit dependence on the target fault which is not presented in Reference [2]. Since the filter is derived similarly to Reference [2], many properties obtained in Reference [2] also apply to this filter. However, some new important properties are given. For example, since the target fault direction is now explicitly in the filter gain calculation, a mechanism is provided which enhances the sensitivity of the filter to the target fault. Furthermore, the projector, which annihilates the residual direction associated with the nuisance faults and is assumed in the problem formulation of Reference [2], is not required in the derivation of this filter. Finally, it is shown that this filter completely blocks the nuisance faults in the limit where the weighting on the nuisance faults is zero. For time-invariant systems, the nuisance faults are placed in a minimal (C, A)-unobservability subspace, and the generalized least-squares fault detection filter becomes equivalent to the unknown input observer. For time-varying systems, the nuisance faults are placed in a similar invariant subspace, and the generalized least-squares fault detection filter extends the unknown input observer to the time-varying case. In the limit, a reduced-order filter is derived for time-varying systems.

The problem is formulated in Section 2 and its solution is derived in Section 3 [2,4]. In Section 4, the filter is derived in the limit [2,5]. In Section 5, it is shown that, in the limit, the nuisance faults are placed in an invariant subspace. In Section 6, the reduced-order filter is derived in the limit. In Section 7, numerical examples are given.

2. PROBLEM FORMULATION

Consider a linear, observable system with two failure modes [1,2]

$$\dot{x} = Ax + Bu + F_1\mu_1 + F_2\mu_2 \tag{1a}$$

$$y = Cx + v \tag{1b}$$

where u is the control input, y is the measurement, v is the sensor noise, μ_1 is the target fault, and μ_2 is the nuisance fault. All system variables belong to real vector spaces, $x \in \mathscr{X}$, $u \in \mathscr{U}$, and $y \in \mathscr{Y}$. System matrices A, B, C, F_1 and F_2 are time-varying and continuously differentiable. The failure modes, μ_1 and μ_2, model the time-varying amplitude of the failure while the failure signatures, F_1 and F_2, model the directional characteristics of a failure. Assume F_1 and F_2 are monic so that $F_1 \neq 0$ and $F_2 \neq 0$ imply $F_1\mu_1 \neq 0$ and $F_2\mu_2 \neq 0$, respectively. In References [1,2], it is shown that this model, used to determine the fault detection filter, represents actuator, sensor and plant faults. There are two assumptions about the system (1) that are needed in order to have a well-conditioned unknown input observer. Assumption 2.1 ensures that the target fault can be isolated from the nuisance fault [1,2]. The output separability test is discussed in Remark 1 of Section 5. Assumption 2.2. ensures a non-zero residual in steady-state when the target fault occurs for time-invariant systems [3,6].

Assumption 2.1.

F_1 and F_2 are output separable.

Assumption 2.2.

For time-invariant systems, (C, A, F_1) does not have invariant zero at origin.

The objective of blocking the nuisance fault while detecting the target fault can be achieved by solving the following min–max problem:

$$\min_{\mu_1} \max_{\mu_2} \max_{x(t_0)} \frac{1}{2}\int_{t_0}^{t} (\|\mu_1\|^2_{Q_1^{-1}} - \|\mu_2\|^2_{\gamma Q_2^{-1}} - \|y - Cx\|^2_{V^{-1}})\,\mathrm{d}\tau - \frac{1}{2}\|x(t_0) - \hat{x}_0\|^2_{\Pi_0} \tag{2}$$

subject to (1a). Note that, without the minimization with respect to μ_1, (2) reduces to the standard least-squares derivation of the Kalman filter [4]. t is the current time and y is assumed given. Q_1, Q_2, V and Π_0 are positive definite. γ is a non-negative scalar. Note that Q_1, Q_2, Π_0 and γ are design parameters to be chosen while V may be physically related to the power spectral density of the sensor noise because of (1b) [4]. The interpretation of the min–max problem is the following. Let μ_1^*, μ_2^* and $x^*(t_0)$ be the optimal strategies for μ_1, μ_2 and $x(t_0)$, respectively. Then, $x^*(\tau \mid Y_t)$, the x associated with μ_1^*, μ_2^* and $x^*(t_0)$, is the optimal trajectory for x where $\tau \in [t_0, t]$ and given the measurement history $Y_t = \{y(\tau) \mid t_0 \leqslant \tau \leqslant t\}$. Since μ_1 maximizes $y - Cx$ and μ_2 minimizes $y - Cx$, $y - Cx^*$ is made primarily sensitive to μ_1 and minimally sensitive to μ_2. However, since x^* is the smoothed estimate of the state, a filtered estimate of the state, called $\hat{x}$, is needed for implementation. From the boundary condition in Section 3, at the current time t, $x^*(t \mid Y_t) = \hat{x}(t)$. Therefore, $y - C\hat{x}$ is primarily sensitive to the target fault and minimally sensitive to the nuisance fault. Note that when Q_1 is larger, $y - C\hat{x}$ is more sensitive to the target fault. When γ is smaller, $y - C\hat{x}$ is less sensitive to the nuisance fault. In Reference [2], the differential game blocks the nuisance fault, but does not enhance the sensitivity to the target fault. In Section 5, it is shown that the filter completely blocks the nuisance fault when γ is zero by placing it into an invariant subspace, called Ker S. Therefore, the residual used for detecting the target fault is

$$r = \hat{H}(y - C\hat{x}) \tag{3}$$

where $\hat{x}$, the filtered estimate of the state, is given in Section 3 and

$$\hat{H}: \mathscr{Y} \to \mathscr{Y}, \qquad \operatorname{Ker} \hat{H} = C \operatorname{Ker} S, \quad \hat{H} = I - C \operatorname{Ker} S[(C \operatorname{Ker} S)^{\mathrm{T}} C \operatorname{Ker} S]^{-1} (C \operatorname{Ker} S)^{\mathrm{T}} \tag{4}$$

Ker S is given and discussed in Sections 4 and 5.

3. SOLUTION

In this section, the min–max problem given by (2) is solved [2,4]. The variational Hamiltonian of the problem is

$$\mathscr{H} = \tfrac{1}{2}(\|\mu_1\|^2_{Q_1^{-1}} - \|\mu_2\|^2_{\gamma Q_2^{-1}} - \|y - Cx\|^2_{V^{-1}}) + \lambda^{\mathrm{T}}(Ax + Bu + F_1\mu_1 + F_2\mu_2)$$

where $\lambda \in \mathscr{R}^n$ is a continuously differentiable Lagrange multiplier. The first-order necessary conditions [4] imply that the optimal strategies for μ_1, μ_2 and the dynamics for λ are

$$\mu_1^* = -Q_1 F_1^{\mathrm{T}} \lambda, \quad \mu_2^* = \frac{1}{\gamma} Q_2 F_2^{\mathrm{T}} \lambda, \quad \dot{\lambda} = -A^{\mathrm{T}}\lambda - C^{\mathrm{T}} V^{-1}(y - Cx)$$

with boundary conditions

$$\lambda(t_0) = \Pi_0[x^*(t_0) - \hat{x}_0], \quad \lambda(t) = 0 \tag{5}$$

By substituting μ_1^* and μ_2^* into (1a), the two-point boundary value problem requires the solution to

$$\begin{bmatrix} \dot{x}^* \\ \dot{\lambda} \end{bmatrix} = \begin{bmatrix} A & \frac{1}{\gamma} F_2 Q_2 F_2^{\mathrm{T}} - F_1 Q_1 F_1^{\mathrm{T}} \\ C^{\mathrm{T}} V^{-1} C & -A^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} x^* \\ \lambda \end{bmatrix} + \begin{bmatrix} Bu \\ -C^{\mathrm{T}} V^{-1} y \end{bmatrix} \tag{6}$$

with boundary conditions (5). The form of (5) suggests that

$$\lambda = \Pi(x^* - \hat{x}) \tag{7}$$

where $\Pi(t_0) = \Pi_0$, $\hat{x}(t_0) = \hat{x}_0$ and $\hat{x}$ is an intermediate state. By differentiating (7), using (6), adding and subtracting $\Pi A \hat{x}$ and $C^{\mathrm{T}} V^{-1} C \hat{x}$, the following dynamic filter structure results:

$$\Pi \dot{\hat{x}} = \Pi A \hat{x} + \Pi B u + C^{\mathrm{T}} V^{-1} (y - C\hat{x}), \quad \hat{x}(t_0) = \hat{x}_0 \tag{8}$$

$$-\dot{\Pi} = \Pi A + A^{\mathrm{T}} \Pi + \Pi \left(\frac{1}{\gamma} F_2 Q_2 F_2^{\mathrm{T}} - F_1 Q_1 F_1^{\mathrm{T}} \right) \Pi - C^{\mathrm{T}} V^{-1} C, \quad \Pi(t_0) = \Pi_0 \tag{9}$$

Since $x^* = \hat{x}$ at current time t (5), the generalized least-squares fault detection filter is (8). Note that (8) is used by the residual (3) to detect the target fault.

4. LIMITING CASE

In this section, the min–max problem (2) is solved in the limit where γ is zero [2,5]. When γ is zero, there is no constraint on μ_2 to minimize $y - Cx$. Therefore, the nuisance fault is completely blocked from the residual which is shown in Section 5.

In the limit, the min–max problem (2) becomes

$$\min_{\mu_1} \max_{\mu_2} \max_{x(t_0)} \frac{1}{2} \int_{t_0}^{t} (\|\mu_1\|^2_{Q_1^{-1}} - \|y - Cx\|^2_{V^{-1}}) \, \mathrm{d}\tau - \frac{1}{2} \|x(t_0) - \hat{x}_0\|^2_{\Pi_0} \tag{10}$$

This problem is singular with respect to μ_2. Therefore, the Goh transformation [5] is used to form a non-singular problem. Let

$$\phi_1(\tau) = \int_{t_0}^{\tau} \mu_2(s) \, \mathrm{d}s, \quad \alpha_1 = x - F_2 \phi_1$$

By differentiating α_1 and using (1a),

$$\dot{\alpha}_1 = A\alpha_1 + Bu + F_1 \mu_1 + B_1 \phi_1 \tag{11}$$

where $B_1 = AF_2 - \dot{F}_2$. By substituting α_1 into (10), the new min–max problem is

$$\min_{\mu_1} \max_{\phi_1} \max_{x_1(t_0^+)} \frac{1}{2}\int_{t_0}^{t} [\|\mu_1\|^2_{Q_1^{-1}} - \|\phi_1\|^2_{F_2^{\mathrm{T}} C^{\mathrm{T}} V^{-1} C F_2} - \|y - C\alpha_1\|^2_{V^{-1}} + (y - C\alpha_1)^{\mathrm{T}} V^{-1} C F_2 \phi_1$$

$$+ \phi_1^{\mathrm{T}} F_2^{\mathrm{T}} C^{\mathrm{T}} V^{-1}(y - C\alpha_1)] \, \mathrm{d}\tau - \frac{1}{2}\|\alpha_1(t_0^+) + F_2\phi_1(t_0^+) - \hat{x}_0\|^2_{\Pi_0} \tag{12}$$

subject to (11). If $F_2^{\mathrm{T}} C^{\mathrm{T}} V^{-1} C F_2$ fails to be positive definite, (12) is still a singular problem with respect to ϕ_1. Then, the Goh transformation has to be used until the problem becomes non-singular. If $F_2^{\mathrm{T}} C^{\mathrm{T}} V^{-1} C F_2 = 0$, let

$$\phi_2(\tau) = \int_{t_0}^{\tau} \phi_1(s) \, \mathrm{d}s, \quad \alpha_2 = \alpha_1 - B_1\phi_2$$

Then, $\dot{\alpha}_2 = A\alpha_2 + Bu + F_1\mu_1 + B_2\phi_2$ where $B_2 = AB_1 - \dot{B}_1$. If $F_2^{\mathrm{T}} C^{\mathrm{T}} V^{-1} C F_2 \geqslant 0$, the Goh transformation is applied only on the singular part [6]. The transformation process stops if the weighting on ϕ_2, $B_1^{\mathrm{T}} C^{\mathrm{T}} V^{-1} C B_1$, is positive definite. Otherwise, continue the transformation until there exists B_k such that the weighting on ϕ_k, $B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1} C B_{k-1}$, is positive definite. Then, in the limit, the min–max problem (2) becomes

$$\min_{\mu_1} \max_{\phi_k} \max_{x_k(t_0^+)} \frac{1}{2}\int_{t_0}^{t} [\|\mu_1\|^2_{Q_1^{-1}} - \|\phi_k\|^2_{B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1} C B_{k-1}} - \|y - C\alpha_k\|^2_{V^{-1}} + (y - C\alpha_k)^{\mathrm{T}} V^{-1} C B_{k-1} \phi_k$$

$$+ \phi_k^{\mathrm{T}} B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1}(y - C\alpha_k)] \, \mathrm{d}\tau - \frac{1}{2}\|\alpha_k(t_0^+) + \bar{B}\bar{\phi}(t_0^+) - \hat{x}_0\|^2_{\Pi_0} \tag{13}$$

subject to $\dot{\alpha}_k = A\alpha_k + Bu + F_1\mu_1 + B_k\phi_k$ where $\bar{B} = [F_2 \; B_1 \; B_2 \; \cdots \; B_{k-1}]$ and $\bar{\phi} = [\phi_1^{\mathrm{T}} \; \phi_2^{\mathrm{T}} \; \cdots \; \phi_k^{\mathrm{T}}]^{\mathrm{T}}$. The min–max problem (13) can be solved similarly to (2). Therefore, the derivation [6] is not repeated here. The limiting generalized least-squares fault detection filter is

$$S\dot{\hat{x}} = SA\hat{x} + SBu + [SB_k(B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1} C B_{k-1})^{-1} B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1} + C^{\mathrm{T}} \bar{H}^{\mathrm{T}} V^{-1} \bar{H}](y - C\hat{x}) \tag{14}$$

where

$$-\dot{S} = S\bar{A} + \bar{A}^{\mathrm{T}} S + S[B_k(B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1} C B_{k-1})^{-1} B_k^{\mathrm{T}} - F_1 Q_1 F_1^{\mathrm{T}}]S - C^{\mathrm{T}} \bar{H}^{\mathrm{T}} V^{-1} \bar{H} C \tag{15}$$

$\bar{H} = I - CB_{k-1}(B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1} C B_{k-1})^{-1} B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1}$ and $\bar{A} = A - B_k(B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1} C B_{k-1})^{-1} B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1} C$ subject to $\hat{x}(t_0^+) = \hat{x}_0$ and $S(t_0^+) = \Pi_0 - \Pi_0 \bar{B}(\bar{B}^{\mathrm{T}} \Pi_0 \bar{B})^{-1} \bar{B}^{\mathrm{T}} \Pi_0$. However, (14) cannot be used because S has a null space which is shown in Theorem 4.1. Therefore, a reduced-order filter for (14) is derived in Section 6.

Theorem 4.1.

$S[B_{k-1} \; B_{k-2} \; \cdots \; B_1 \; F_2] = 0.$

Proof. The proof is similar to Reference [2] and can be found in Reference [6]. □

5. PROPERTIES OF THE NULL SPACE OF S

In this section, some properties of the null space of S are given. It is shown that the null space of S is equivalent to the minimal (C, A) – unobservability subspace for time-invariant systems and a similar invariant subspace for time-varying systems. Therefore, the limiting generalized least-squares fault detection filter is equivalent to the unknown input observer and extends it to the time-varying case. The minimal (C, A)-unobservability subspace is a subspace which is $(A - LC)$-invariant and unobservable with respect to $(\tilde{H}C, A - LC)$ for some filter gain L and projector $\tilde{H}$ [1]. One method for computing the minimal (C, A)-unobservability subspace of F_2, called $\mathscr{T}_2$ here, is $\mathscr{T}_2 = \mathscr{W}_2 \oplus \mathscr{V}_2$ [1] where $\mathscr{W}_2 = [B_{k-1} \;\; B_{k-2} \;\; \cdots \;\; B_1 \;\; F_2]$ is the minimal (C, A)-invariant subspace of F_2 and $\mathscr{V}_2$ is the subspace spanned by the invariant zero directions of (C, A, F_2). Note that the associated $\tilde{H}$ is

$$\tilde{H}: \mathscr{Y} \to \mathscr{Y}, \quad \operatorname{Ker} \tilde{H} = CB_{k-1}, \quad \tilde{H} = I - CB_{k-1}[(CB_{k-1})^{\mathrm{T}} CB_{k-1}]^{-1} (CB_{k-1})^{\mathrm{T}} \tag{16}$$

Note that $\operatorname{Ker} \tilde{H} = \operatorname{Ker} \hat{H}$.

Theorem 5.1 shows that the null space of S is a (C, A)-invariant subspace. Theorem 5.2 shows that the null space of S is contained in the unobservable subspace of $(\hat{H}C, A - LC)$.

Theorem 5.1.

Ker S is a (C, A)-invariant subspace.

Proof. The dynamic equation of the error, $e = x - \hat{x}$, in the absence of the target fault and sensor noise can be obtained by using (1) and (14):

$$S\dot{e} = [SA + SB_k(B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1} CB_{k-1})^{-1} B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1} C + C^{\mathrm{T}} \hat{H}^{\mathrm{T}} V^{-1} \hat{H} C] e$$

because $SF_2 = 0$. By adding $\dot{S}e$ to both sides and using (15),

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\tau}(Se) = & -\{[A - B_k(B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1} CB_{k-1})^{-1} B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1} C]^{\mathrm{T}} \\ & + S[-F_1 Q_1 F_1^{\mathrm{T}} + B_k(B_{k-1}^{\mathrm{T}} C^{\mathrm{T}} V^{-1} CB_{k-1})^{-1} B_k^{\mathrm{T}}]\} Se \end{aligned} \tag{17}$$

If the error initially lies in Ker S, (17) implies that the error will never leave Ker S. Therefore, Ker S is a (C, A)-invariant subspace. □

Theorem 5.2.

Ker S is contained in the unobservable subspace of $(\hat{H}C, A - LC)$.

Proof. Let $\zeta \in$ Ker S. By multiplying (15) by ζ^{T} from the left and ζ from the right,

$$\frac{\mathrm{d}}{\mathrm{d}\tau}(\zeta^{\mathrm{T}} S \zeta) = \zeta^{\mathrm{T}} C^{\mathrm{T}} \hat{H}^{\mathrm{T}} V^{-1} \hat{H} C \zeta = 0$$

Then, $\bar{H}C\zeta = 0$ because $\hat{H}C\zeta = 0$ and $\operatorname{Ker}\hat{H} = \operatorname{Ker}\bar{H}$. From Theorem 5.1, $\operatorname{Ker} S$ is a (C, A)-invariant subspace. Therefore, $\operatorname{Ker} S$ is contained in the unobservable subspace of $(\hat{H}C, A - LC)$. □

From Theorem 4.1, $C \operatorname{Ker} S \supseteq CB_{k-1}$. From Theorem 5.2, $C \operatorname{Ker} S \subseteq CB_{k-1}$. Therefore, $C \operatorname{Ker} S = CB_{k-1}$ and $\hat{H}$ (4) is equivalent to $\bar{H}$ (16). Note that (16) is a better way to form $\hat{H}$ which is used by the residual (3) because it does not require the solution to the limiting Riccati Equation (15).

For time-invariant systems, it is important to discuss the invariant zero directions when designing the fault detection filter. The invariant zeros of (C, A, F_2) will become part of the eigenvalues of the filter if their associated invariant zero directions are not included in the invariant subspace of F_2 [1]. From Reference [3,6], the null space of S includes all the invariant zero directions if the nuisance fault direction is modified to the invariant zero directions. Therefore, the invariant zeros will not become part of the filter eigenvalues. From Theorem 4.1 and modified nuisance fault direction, the null space of S contains the minimal (C, A)-unobservability subspace of F_2. By combining with Theorem 5.2, the null space of S is equivalent to the minimal (C, A)-unobservability subspace of F_2, and the limiting generalized least-squares fault detection filter is equivalent to the unknown input observer. Note that the invariant zero and minimal (C, A)-unobservability subspace are only defined for time-invariant systems. For time-varying systems, Theorems 4.1, 5.1 and 5.2 imply that the null space of S is a similar invariant subspace.

Remark 1.

In order to detect the target fault, F_1 cannot intersect the null space of S which is unobservable to the residual. If it does, the target fault will be difficult or impossible to detect even though the filter can still be derived by solving the min–max problem. If F_1 does not intersect the null space of S, F_1 and F_2 are called output separable [1], and the output separability test can be stated as $CB_{k-1} \cap C\tilde{B}_{k-1} = \emptyset$ where $\tilde{B}_{k-1}$ is the Goh transformation of F_1.

6. REDUCED-ORDER FILTER

In this section, the reduced-order filter is derived for the limiting generalized least-squares fault detection filter (14). The reduced-order filter is necessary for implementation because (14) cannot be used due to the null space of S. Since S is non-negative definite, there exists a state transformation Γ such that

$$\Gamma^{\mathrm{T}} S \Gamma = \begin{bmatrix} \bar{S} & 0 \\ 0 & 0 \end{bmatrix} \tag{18}$$

where $\bar{S}$ is positive definite. Theorem 6.1 provides a way to form the transformation.

Theorem 6.1.

There exists a state transformation Γ where

$$[Z \ \ \operatorname{Ker} S] = \Gamma \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \tag{19}$$

Z is any $n \times (n - k_2)$ continuously differentiable matrix such that itself and Ker S span the state space where $n = \dim \mathcal{X}$ and $k_2 = \dim(\text{Ker } S)$. Z_1 and Z_2 are any $(n - k_2) \times (n - k_2)$ and $k_2 \times k_2$ invertible continuously differentiable matrices, respectively. Then, the Γ obtained from (19) satisfies (18).

Proof.

$$\text{Ker } S = \Gamma \begin{bmatrix} 0 \\ Z_2 \end{bmatrix} \Rightarrow S\Gamma \begin{bmatrix} 0 \\ Z_2 \end{bmatrix} = 0 \Rightarrow \Gamma^{\mathrm{T}} S \Gamma \begin{bmatrix} 0 \\ Z_2 \end{bmatrix} = 0$$

Since Z_2 is invertible by definition and $\Gamma^{\mathrm{T}} S \Gamma$ is symmetric, (18) is true. □

Note that Theorem 6.1 does not define Γ uniquely and Γ can be computed *a priori* because Ker S can be obtained *a priori*.

By applying the transformation to the estimator state, $\Gamma^{-1}\hat{x} \triangleq \hat{\eta} = [\hat{\eta}_1^{\mathrm{T}} \ \hat{\eta}_2^{\mathrm{T}}]^{\mathrm{T}}$. By multiplying (14) by Γ^{T} from the left, using $\Gamma\Gamma^{-1} = I$, and adding $\Gamma^{\mathrm{T}} S \Gamma \dot{\Gamma}^{-1} \hat{x}$ to both sides, the limiting filter can be transformed into two equations,

$$\begin{aligned} \bar{S}\dot{\hat{\eta}}_1 &= \bar{S}(A_{11} - \Gamma_{11})\hat{\eta}_1 + \bar{S}(A_{12} - \Gamma_{12})\hat{\eta}_2 + \bar{S}M_1 u \\ &\quad + [\bar{S}G_1(D_2^{\mathrm{T}} C_2^{\mathrm{T}} V^{-1} C_2 D_2)^{-1} D_2^{\mathrm{T}} C_2^{\mathrm{T}} V^{-1} + C_1^{\mathrm{T}} \bar{H}^{\mathrm{T}} V^{-1} \bar{H}](y - C_1\hat{\eta}_1 - C_2\hat{\eta}_2) \end{aligned} \tag{20a}$$

$$0 = C_2^{\mathrm{T}} \bar{H}^{\mathrm{T}} V^{-1} \bar{H}(y - C_1\hat{\eta}_1 - C_2\hat{\eta}_2) \tag{20b}$$

where

$$\Gamma^{-1}\dot{\Gamma} = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}, \quad \Gamma^{-1}A\Gamma = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad \Gamma^{-1}B = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}, \quad C\Gamma = [C_1 \ \ C_2]$$

$$\Gamma^{-1}F_1 = \begin{bmatrix} N_1 \\ N_2 \end{bmatrix}, \quad \Gamma^{-1}B_{k-1} = \begin{bmatrix} 0 \\ D_2 \end{bmatrix}, \quad \Gamma^{-1}B_k = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$$

Note that Γ^{-1} and $\dot{\Gamma}$ can be computed *a priori* from (19). From (20b),

$$\bar{H}C_2 = 0 \tag{21}$$

because $y - C_1\hat{\eta}_1 - C_2\hat{\eta}_2$ is arbitrary. By multiplying (15) by Γ^{T} from the left and Γ from the right, subtracting $\dot{\Gamma}^{\mathrm{T}} S \Gamma$ and $\Gamma S \dot{\Gamma}^{\mathrm{T}}$ from both sides, and using $\Gamma\Gamma^{-1} = I$, the limiting Riccati equation can be transformed into two equations,

$$0 = \bar{S}[A_{12} - \Gamma_{12} - G_1(D_2^{\mathrm{T}} C_2^{\mathrm{T}} V^{-1} C_2 D_2)^{-1} D_2^{\mathrm{T}} C_2^{\mathrm{T}} V^{-1} C_2] \tag{22}$$

$$-\dot{\bar{S}} = \bar{S}[A_{11} - \Gamma_{11} - G_1(D_2^{\mathrm{T}}C_2^{\mathrm{T}}V^{-1}C_2D_2)^{-1}D_2^{\mathrm{T}}C_2^{\mathrm{T}}V^{-1}C_1]$$
$$+ [A_{11} - \Gamma_{11} - G_1(D_2^{\mathrm{T}}C_2^{\mathrm{T}}V^{-1}C_2D_2)^{-1}D_2^{\mathrm{T}}C_2^{\mathrm{T}}V^{-1}C_1]^{\mathrm{T}}\bar{S}$$
$$+ \bar{S}[-N_1Q_1N_1^{\mathrm{T}} + G_1(D_2^{\mathrm{T}}C_2^{\mathrm{T}}V^{-1}C_2D_2)^{-1}G_1^{\mathrm{T}}]\bar{S} - C_1^{\mathrm{T}}\bar{H}^{\mathrm{T}}V^{-1}\bar{H}C_1 \qquad (23)$$

By substituting (21) and (22) into (20a), the reduced-order limiting generalized least-squares fault detection filter is

$$\dot{\hat{\eta}}_1 = (A_{11} - \Gamma_{11})\hat{\eta}_1 + M_1u + [G_1(D_2^{\mathrm{T}}C_2^{\mathrm{T}}V^{-1}C_2D_2)^{-1}D_2^{\mathrm{T}}C_2^{\mathrm{T}}V^{-1} + \bar{S}^{-1}C_1^{\mathrm{T}}\bar{H}^{\mathrm{T}}V^{-1}\bar{H}](y - C_1\hat{\eta}_1) \qquad (24)$$

Note that Γ_{11} can be computed *a priori*. In the limit, the residual (3) becomes

$$r = \hat{H}(y - C_1\hat{\eta}_1) \qquad (25)$$

because $\hat{H}C_2 = 0$ from (21) and Ker $\hat{H}$ = Ker $\bar{H}$.

7. EXAMPLE

In this section, two numerical examples are used to demonstrate the performance of the generalized least-squares fault detection filter. In Section 7.1, the filter is applied to a time -invariant system. In Section 7.2, the filter is applied to a time-varying system.

7.1. Example 1

In this section, two cases for a time-invariant problem are presented. The first one shows that the sensitivity of the filter (8) to the nuisance fault decreases when γ is smaller. The second one shows that the sensitivity of the reduced-order limiting filter (24) to the target fault increases when Q_1 is larger. The system matrices are

$$A = \begin{bmatrix} 0 & 3 & 4 \\ 1 & 2 & 3 \\ 0 & 2 & 5 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad F_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad F_2 = \begin{bmatrix} 5 \\ 1 \\ 1 \end{bmatrix}$$

In the first case, the steady-state solutions to the Riccati equation (9) are obtained with weightings chosen as $Q_1 = 1$, $Q_2 = 1$, and $V = I$ when $\gamma = 10^{-4}$ and 10^{-6}, respectively. The top two figures of Figure 1 show the frequency response from both faults to the residual (3). The left one is $\gamma = 10^{-4}$, and the right one is $\gamma = 10^{-6}$. The solid lines represent the target fault, and the dashed lines represent the nuisance fault. This example shows that the nuisance fault transmission can be reduced by using a smaller γ while the target fault transmission is not affected.

In the second case, the steady-state solutions to the reduced-order limiting Riccati equation (23) are obtained with $V = 10^{-4}I$ when $Q_1 = 0$ and 0.0019, respectively. The lower two figures of Figure 1 show the frequency response from the target fault and sensor noise to the residual (25). The left one is $Q_1 = 0$, and the right one is $Q_1 = 0.0019$. The solid lines represent the target fault, and the dashed lines represent the sensor noise. This example shows that the

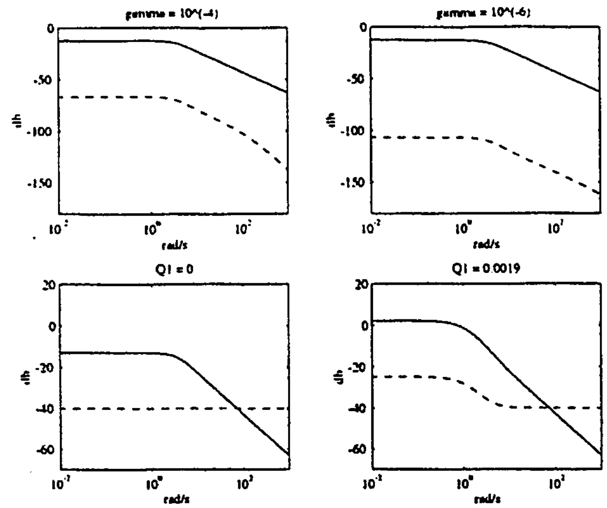

Figure 1. Frequency response of the residual.

sensitivity of the filter to the target fault can be enhanced by using a larger Q_1. The sensor noise transmission also increases because part of the sensor noise comes through the same direction as the target fault. However, the sensor noise transmission is small compared to the target fault transmission. In this case, the nuisance fault transmission stays zero and is not shown in these figures. Note that when $Q_1 = 0$, the generalized least-squares fault detection filter is similar to Reference [2] which does not enhance the target fault transmission.

7.2. *Example 2*

In this section, the filter (8) and the reduced-order limiting filter (24) are applied to a time-varying system which is from modifying the time-invariant system in the previous section by adding some time-varying elements to A and F_2 matrices while C and F_1 matrices are the same:

$$A = \begin{bmatrix} -\cos t & 3 + 2\sin t & 4 \\ 1 & 2 & 3 - 2\cos t \\ 5\sin t & 2 & 5 + 3\cos t \end{bmatrix}, \quad F_2 = \begin{bmatrix} 5 - 2\cos t \\ 1 \\ 1 + \sin t \end{bmatrix}$$

The Riccati equation (9) is solved with $Q_1 = 1$, $Q_2 = 1$, $V = I$ and $\gamma = 10^{-5}$ for $t \in [0, 25]$. The reduced-order limiting Riccati equation (23) is solved with the same Q_1 and V. Figure 2 shows the time response of the norm of the residuals when there is no fault, a target fault and a nuisance fault, respectively. The faults are unit steps that occur at the fifth second. In each case, there is no sensor noise. The left three figures show the residual (3) for the filter (8). There is a small nuisance fault transmission because (8) is an approximate unknown input observer. The right three figures show the residual (25) for the reduced-order limiting filter (24). Note that the nuisance fault transmission is zero. This example shows that both filters, (8) and (24), work well for time-varying systems.

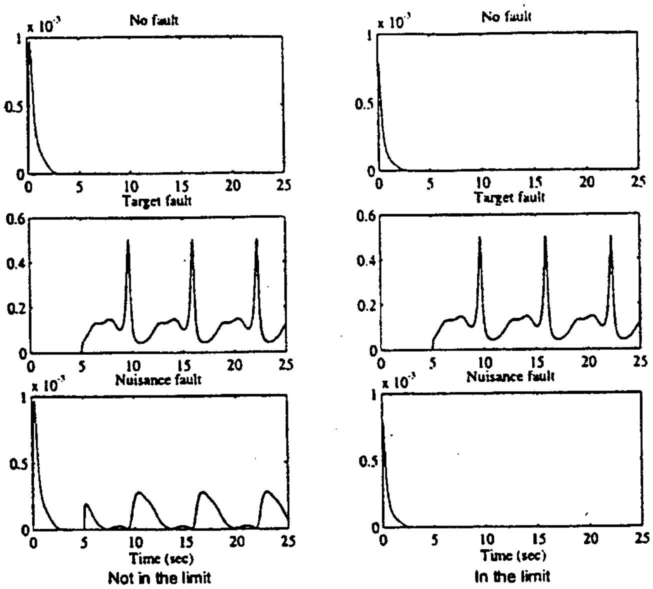

Figure 2. Time response of the residual.

8. CONCLUSION

The generalized least-squares fault detection filter is derived from solving a min–max problem which makes the residual sensitive to the target fault, but insensitive to the nuisance faults. In the limit where the weighting on the nuisance faults is zero, the filter becomes equivalent to the unknown input observer which places the nuisance faults into a minimal (C, A)-unobservability subspace and there exists a reduced-order filter. Since the target fault is explicit in the problem formulation, the sensitivity of the filter to the target fault can be enhanced. Filter designs can be obtained for both linear-time-invariant and time-varying systems.

ACKNOWLEDGEMENTS

This work was sponsored by Air Force Office of Scientific Research, Award No. F49620-97-1-0272 and by the California Department of Transportation, Agreement No. 65A0013, MOU315.

REFERENCES

1. Massoumnia M-A, Verghese GC, Willsky AS. Failure detection and identification. *IEEE Transactions on Automatic Control* 1989; **AC-34**(3):316–321.
2. Chung WH, Speyer JL. A game theoretic fault detection filter. *IEEE Transactions on Automatic Control* 1998; **AC-43**(2):143–161.
3. Chen RH, Speyer JL. Optimal stochastic fault detection filter. *Proceedings of the American Control Conference*, 1999; 91–96.
4. Bryson AE, Ho Y-C. *Applied Optimal Control: Optimization, Estimation and Control*. Hemisphere: Washington, DC, 1975.
5. Bell DJ, Jacobsen DH. *Singular Optimal Control Problems. Mathematics in Science and Engineering*, vol. 117. Academic Press: New York, 1975.
6. Chen RH. Fault detection filters for robust analytical redundancy. *PhD Thesis*, University of California, Los Angeles, 2000.

APPENDIX E

Optimal Stochastic Multiple-Fault Detection Filter

R.H. Chen and J.L. Speyer

Proceedings of the 38th
Conference on Decision & Control
Phoenix, Arizona USA • December 1999

Optimal Stochastic Multiple-Fault Detection Filter [1]

Robert H. Chen and Jason L. Speyer
Mechanical and Aerospace Engineering Department
University of California, Los Angeles, California 90095-1597
chrobert@talus.seas.ucla.edu, speyer@seas.ucla.edu

Abstract

A class of robust fault detection filters is generalized from detecting single fault to multiple faults. This generalization is called the optimal stochastic multiple-fault detection filter since in the formulation, the unknown fault amplitudes are modeled as white noise. The residual space of the filter is divided into several subspaces and each subspace is sensitive to only one fault (target fault), but not to other faults (nuisance faults), in the sense that the transmission from nuisance faults to the target residual space is small while the transmission from target fault is large. It is shown that this filter approximates the properties of the classical fault detection filter such that in the limit where the nuisance fault weighting goes to infinity, the optimal stochastic multiple-fault detection filter is equivalent to the Beard-Jones fault detection filter when there is no complementary subspace. A numerical example also shows that this filter is an approximate Beard-Jones fault detection filter even when complementary subspace exists. This filter combines the advantages of the robust single-fault detection filter and Beard-Jones fault detection filter.

1 Introduction

Any system under automatic control demands a high degree of system reliability and this requires a health monitoring system capable of detecting any system, actuator and sensor fault as it occurs and identifying the faulty component. One approach, analytical redundancy, uses the modeled dynamic relationship between system inputs and measured system outputs to form a residual process used for detecting and identifying faults. Nominally, the residual is nonzero only when a fault has occurred and is zero at other times.

A popular approach to analytical redundancy is the detection filter which was first introduced by [1] and refined by [2]. It is also known as the Beard-Jones fault detection filter. A geometric interpretation and a spectral approach of this filter are given in [3] and [4], respectively. Design algorithms have been developed [5, 6] which improved detection filter robustness. The idea of a detection filter is to put the reachable subspace of each fault into invariant subspaces which do not overlap with each other. Then, when a nonzero residual is detected, a fault can be announced and identified by projecting the residual onto each of the invariant subspaces. Therefore, multiple faults can be monitored in one filter.

Another related approach, the unknown input observer [7], simplifies the detection filter problem by dividing the faults into a target fault and nuisance fault group where the nuisance faults are placed into one invariant subspace. Although only one fault can be detected in each unknown input observer, additional flexibility in fault detection filter design for robustness and time-varying system is obtained by using an approximate fault detection filter [8, 9, 10, 11, 12].

In this paper, an extension of the optimal stochastic fault detection filter [11] is presented. The optimal stochastic fault detection filter, which is an approximate unknown input observer, allows additional robustness in the fault detection filter design. However, it can detect only one fault in each filter. In contrast, the Beard-Jones fault detection filter can detect multiple faults in one filter, but is not very robust. From the problem formulation of the optimal stochastic fault detection filter, it seems natural that the multiple faults objective may be achieved. This is done by dividing the residual space of the filter into several subspaces by projectors and having each subspace sensitive to only one fault (target fault), but not to other faults (nuisance faults), in the sense that the transmission from nuisance faults to the target residual space is small while the transmission from target fault is large. In the limit where the nuisance fault weighting goes to infinity and in the absence of sensor noise, it is shown that the optimal stochastic multiple-fault detection filter becomes a Beard-Jones fault detection filter when there is no complementary subspace. Note that the $\mathcal{H}_\infty$ bounded fault detection filter [6] imposed the detection filter structure constraint while the detection filter structure is generated from the problem formulation of the optimal stochastic multiple-fault detection filter. Also, a numerical example shows that this filter is an approxi-

[1] This work was sponsored by Air Force Office of Scientific Research, Award No. F49620-97-1-0272 and NASA-Ames Research Center, Cooperative Agreement NCC2-374, Supplement 19

mate Beard-Jones fault detection filter when it is not in the limit even with the existence of the complementary subspace.

The problem is formulated in Section 2 and the solution is derived in Section 3. In Section 4, the filter is derived for the limiting case when there is no complementary subspace. In Section 5, a numerical example is given.

2 Problem Formulation

In this section, the fault detection filter problem is formulated. From [1, 3, 4, 8], a linear time-invariant, (C, A) observable system with q plant, actuator and sensor faults can be modeled by

$$\dot{x} = Ax + Bu + \sum_{i=1}^{q} F_i \mu_i \tag{1a}$$

$$y = Cx + v \tag{1b}$$

where u is control input, y is measurement and v is sensor noise. The failure modes μ_i are vectors that are unknown and arbitrary functions of time and are zero when there is no failure. The failure signatures F_i are maps that are known. A failure mode μ_i models the time-varying amplitude of a failure while a failure signature F_i models the directional characteristics of a failure. Assume the F_i are monic so that $\mu_i \neq 0$ implies $F_i\mu_i \neq 0$.

There are two assumptions about system (1) in order have a well-conditioned fault detection filter. Assumption 2.1 ensures the separation of faults $\mu_i, i = 1, \cdots, q$ [3, 8]. Assumption 2.2 ensures a nonzero residual in steady state when the target fault occurs [11].

Assumption 2.1. $F_1, \cdots, F_q$ are output separable.

Assumption 2.2. $(C, A, F_i), i=1,\cdots,q$, do not have transmission zeros at origin.

Assume $\mu_i, i = 1, \cdots, q$, and v are zero mean, white Gaussian noise with

$$E[\mu_i(t)\mu_j(\tau)^T] = \begin{cases} Q_i\delta(t-\tau) & , \quad i = j \\ 0 & , \quad i \neq j \end{cases} \tag{2a}$$

$$E[v(t)v(\tau)^T] = V\delta(t-\tau) \tag{2b}$$

and $E[x(t_0)x(t_0)^T] = P_0$. Also, $\mu_i, i = 1, \cdots, q$, and v are uncorrelated with each other and with $x(t_0)$. For simplicity, the following notation is made for use later.

$$\hat{\mu}_i = [\ \mu_1 \ \cdots \ \mu_{i-1} \ \ \mu_{i+1} \ \cdots \ \mu_q\ ]$$
$$\hat{F}_i = [\ F_1 \ \cdots \ F_{i-1} \ \ F_{i+1} \ \cdots \ F_q\ ]$$
$$\hat{Q}_i = E[\hat{\mu}_i(t)\hat{\mu}_i(t)^T]$$

The objective of the optimal stochastic multiple-fault detection filter problem is to find a filter gain L for the linear observer,

$$\dot{\hat{x}} = A\hat{x} + Bu + L(y - C\hat{x})$$

and the residual,

$$r = y - C\hat{x} \tag{3}$$

such that each projected residual $\hat{H}_i r$ is affected essentially only by its target fault μ_i, and minimally by its nuisance fault $\hat{\mu}_i$, sensor noise v and initial condition error $x(t_0) - \hat{x}(t_0)$. $\hat{H}_i$ are projectors also used by the Beard-Jones fault detection filter which map the reachable subspace of $\hat{\mu}_i$ to zero.

$$\hat{H}_i : \mathcal{Y} \to \mathcal{Y} \ , \ \operatorname{Ker} \hat{H}_i = C\hat{\mathcal{T}}_i$$

where $\hat{\mathcal{T}}_i$ is the minimal (C, A)-unobservability subspace of $\hat{F}_i$ with $\hat{k}_i = \dim \hat{\mathcal{T}}_i$. A minimal (C, A)-unobservability subspace [3, 7] implies that there is a projector $\hat{H}$ induced from the fault directions such that $(\hat{H}C, A - LC)$ has an unobservable subspace for some filter gain L. The error, $e = x - \hat{x}$, can be written as

$$e(t) = \Phi(t, t_0)e(t_0) + \int_{t_0}^{t} \Phi(t, \tau)\left(\sum_{i=1}^{q} F_i\mu_i - Lv\right) d\tau \tag{4}$$

subject to

$$\frac{d}{dt}\Phi(t, t_0) = (A - LC)\Phi(t, t_0) \ , \quad \Phi(t_0, t_0) = I \tag{5}$$

And the residual (3) becomes $r = Ce + v$.

Now a performance index is needed for deriving the filter gain L. It seems that the most natural choice is to have the performance index be associated with the residual (i.e., $\hat{H}_i(Ce + v)$). However, it is unusable from statistical viewpoint since the variance of the residual generates a δ-function due to the sensor noise. The next choice is for the performance index to be associated with the output space $\mathcal{Y}$ (i.e., $\hat{H}_iCe$). However, a unique solution can not be obtained from minimizing the performance index associated with the output space because the information on the null space of C is not available. This will become clear in Section 3. Therefore, the performance index will be associated with the state space $\mathcal{X}$ (i.e., H_ie) which means the influence of $\hat{\mu}_i$, v and $e(t_0)$ on H_ie is minimized while the influence of μ_i is maximized. The H_i associated with $\hat{H}_i$ is

$$H_i : \mathcal{X} \to \mathcal{X} \ , \ \operatorname{Ker} H_i = \hat{\mathcal{T}}_i \ , \ H_i = I - \hat{\mathcal{T}}_i[\hat{\mathcal{T}}_i^T\hat{\mathcal{T}}_i]^{-1}\hat{\mathcal{T}}_i^T \tag{6}$$

Note that Beard-Jones fault detection filter also works on the state space by assigning the eigenstructures. In Section 4, it will be shown that the projectors (6) will minimize the performance index in the limit.

Assumption 2.3. The invariant zero directions associated with the invariant zeros of (C, A, F_i) on the left-half plane are included with the fault direction F_i to produce the minimal (C, A)-unobservability subspace of F_i, $\mathcal{T}_i$.

Remark 1. From [3], all invariant zero directions of (C, A, F_i) have to be included in $\mathcal{T}_i$ or the invariant zeros will be part of the eigenvalues of the filter. From the approach of [8], the invariant zero directions associated with the invariant zeros on the right-half plane and imaginary axis are automatically included in $\mathcal{T}_i$. From [11, 12], the invariant zero directions associated with the invariant zeros on the left-half plane will also be included in $\mathcal{T}_i$ only if the fault direction F_i is modified. ◀

Define

$$h_i(t) = H_i \int_{t_0}^{t} \Phi(t,\tau) F_i \mu_i d\tau \tag{7a}$$

$$\hat{h}_i(t) = H_i \int_{t_0}^{t} \Phi(t,\tau) \hat{F}_i \hat{\mu}_i d\tau \tag{7b}$$

$$h_{iv}(t) = H_i \left[\Phi(t,t_0)e(t_0) - \int_{t_0}^{t} \Phi(t,\tau) L v d\tau \right] \tag{7c}$$

From (4), h_i represents the transmission from target fault to part of the error $H_i e$. $\hat{h}_i$ represents the transmission from nuisance faults to $H_i e$. h_{iv} represents the transmission from sensor noise and initial condition error to $H_i e$. Since the objective is to pick a filter gain L such that each $H_i e$ is sensitive only to its target fault, but not its nuisance fault, sensor noise and initial condition error, $\hat{h}_i$ and h_{iv} are expected to be small while h_i to be large. This can be formulated as a non-convex minimization problem,

$$\min_L J = \min_L \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \text{tr} \left\{ \frac{1}{\gamma} \text{E} \left[\sum_{i=1}^{q} (\hat{h}_i \hat{h}_i^T + h_{iv} h_{iv}^T) \right] - \text{E} \left[\sum_{i=1}^{q} h_i h_i^T \right] \right\} dt$$

where $\gamma \geq 0$ and t_1 is the final time. The trace operator is used because the variance is a matrix.

3 Solution to the Disturbance Attenuation Problem

By using (2) and (7), the cost J can be written as

$$J = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \text{tr} \left\{ \sum_{i=1}^{q} \left[H_i \int_{t_0}^{t} \Phi(t,\tau) \left(\frac{LVL^T}{\gamma} + \frac{\hat{F}_i \hat{Q}_i \hat{F}_i^T}{\gamma} - F_i Q_i F_i^T \right) \Phi(t,\tau)^T d\tau H_i + H_i \Phi(t,t_0) \frac{P_0}{\gamma} \Phi(t,t_0)^T H_i \right] \right\} dt$$

To put the optimization problem in a more transparent context, J is manipulated by adding zero term

$$\frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \text{tr} \left\{ \sum_{i=1}^{q} H_i \left[\Phi(t,t) P_i(t) \Phi(t,t)^T - \Phi(t,t_0) P_i(t_0) \Phi(t,t_0)^T - \int_{t_0}^{t} \frac{d}{d\tau} (\Phi(t,\tau) P_i \Phi(t,\tau)^T) d\tau \right] H_i \right\} dt$$

Then, the problem can be rewritten as

$$\begin{aligned} \min_L J &= \min_L \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \frac{1}{\gamma} \text{tr} \left\{ \sum_{i=1}^{q} \left[H_i \int_{t_0}^{t} \Phi(t,\tau) (L - \gamma P_i C^T V^{-1}) V (L - \gamma P_i C^T V^{-1})^T \Phi(t,\tau)^T d\tau H_i \right] \right\} dt \\ &= \min_L \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \frac{1}{\gamma} \text{tr} \left(\sum_{i=1}^{q} H_i W_i(t) H_i \right) dt \end{aligned} \tag{8}$$

subject to

$$\dot{P}_i = A P_i + P_i A^T - \gamma P_i C^T V^{-1} C P_i + \frac{\hat{F}_i \hat{Q}_i \hat{F}_i^T}{\gamma} - F_i Q_i F_i^T \tag{9}$$

$$\begin{aligned} \dot{W}_i =& (A - LC) W_i + W_i (A - LC)^T \\ &+ (L - \gamma P_i C^T V^{-1}) V (L - \gamma P_i C^T V^{-1})^T \end{aligned} \tag{10}$$

where $P_i(t_0) = P_0/\gamma$ and $W_i(t_0) = 0$. The term $\frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \frac{1}{\gamma} \text{tr} [\sum_{i=1}^{q} H_i P_i(t) H_i] \, dt$ is dropped because H_i is not being optimalized here but chosen as in (6). See [11] for extension.

The variational Hamiltonian of the problem is

$$\begin{aligned} \mathcal{H} = \text{tr} \Big\{ \sum_{i=1}^{q} \{ & H_i W_i H_i + \mathcal{K}_i [(A - LC) W_i + W_i (A - LC)^T \\ & + (L - \gamma P_i C^T V^{-1}) V (L - \gamma P_i C^T V^{-1})^T] \} \Big\} \end{aligned} \tag{11}$$

where $\mathcal{K}_i(t) \in \mathcal{R}^{n \times n}$ is a continuously differentiable matrix Lagrange multiplier. Note that $P_i, i = 1, \cdots, q$, are independent of L. The first-order necessary conditions imply that the optimal strategy for L and dynamics for $\mathcal{K}_i$ are

$$\frac{\partial \mathcal{H}}{\partial L} = \sum_{i=1}^{q} [-2 C W_i \mathcal{K}_i + 2 V (L^* - \gamma P_i C^T V^{-1})^T \mathcal{K}_i] = 0 \tag{12}$$

$$\Rightarrow L^* = (\sum_{i=1}^{q} \mathcal{K}_i)^{-1} [\sum_{i=1}^{q} \mathcal{K}_i (\gamma P_i + W_i)] C^T V^{-1} \tag{13}$$

$$-\dot{\mathcal{K}}_i = \frac{\partial \mathcal{H}}{\partial W_i} = H_i + \mathcal{K}_i (A - LC) + (A - LC)^T \mathcal{K}_i = 0 \tag{14}$$

where $\mathcal{K}_i(t_1) = 0$. Note that $\mathcal{K}_i = \mathcal{K}_i^T$. Although the optimal filter gain L^*, (13) subject to (10), (14) and (9), requires the solution to a two-point boundary value problem, it can be computed off-line.

More realistically, the infinite-time case allows a time-invariant L^* where $2q$ algebraic Lyapunov equations

(10) and (14) (i.e., $\dot{W}_i = \dot{\mathcal{K}}_i = 0$), coupled by (13), are to be solved. An alternative is to use a gradient method to numerically solve

$$\lim_{t_1 - t_0 \to \infty} \min_L J = \min_L \operatorname{tr}\left(\frac{1}{\gamma}\sum_{i=1}^{q} H_i W_i H_i\right)$$

where W_i is the solution to the algebraic (10).

Remark 2. The stability of the filter depends on the existence of L^*. If there is a L^* such that the cost J is at its minimum, $A - LC$ has to be stable otherwise J will become unbounded. ◀

Remark 3. If the cost J is associated with the output space (i.e., H in $\mathcal{H}$ (11) is replaced by $\hat{H}C$), the Lagrange multiplier $\mathcal{K}_i(t) \in \mathcal{R}^{m\times n}$ and therefore L^* is not unique from (12). ◀

4 Limiting Case

In this section, the limit of the optimal stochastic multiple-fault detection filter is investigated where $V \to 0$ as $\gamma \to 0$ in such a way that $\gamma V^{-1} \to \bar{V}^{-1}$. Note that V has to go to zero as $\gamma \to 0$ because physically sensor noise will produce a nonzero transmission. Similarly, the variance of the initial condition error $P_0 \to 0$ as $\gamma \to 0$ such that $\gamma P_0^{-1} \to \bar{\Pi}_0$. The infinite-time L^* (13) will be simplified for the limiting case and compared to Beard-Jones fault detection filter. Assumption 4.1 is used to simplify L^* and the following analysis.

Assumption 4.1. There is no complementary subspace.

Therefore, $\sum_{i=1}^{q} \mathcal{T}_i$ spans the state space $\mathcal{X}$ where $\mathcal{T}_i$ is the minimal (C, A)-unobservability subspace of F_i. Assume,

$$(A - LC)\mathcal{T}_i \subseteq \mathcal{T}_i \tag{15}$$

for $i = 1, \cdots, q$, which will be shown in Theorem 4.6. Lemmas 4.1 and 4.2 show that $\mathcal{K}_i$ has similar properties to H_i.

Lemma 4.1. $\mathcal{K}_i \hat{\mathcal{T}}_i = 0$.

Proof. By multiplying algebraic form of (14) by $\hat{v}_{ij}$ from the right,

$$\mathcal{K}_i(A - LC)\hat{v}_{ij} + (A - LC)^T \mathcal{K}_i \hat{v}_{ij} = 0 \tag{16}$$

From (15), let $\hat{v}_{ij}$, $j = 1, \cdots, \hat{k}_i$, span $\hat{\mathcal{T}}_i$ such that $(A - LC)\hat{v}_{ij} = \hat{\sigma}_{ij}\hat{v}_{ij}$. Then, (16) becomes

$$[\hat{\sigma}_{ij} I + (A - LC)^T]\mathcal{K}_i \hat{v}_{ij} = 0$$

which implies $\mathcal{K}_i \hat{v}_{ij} = 0$ because $A - LC$ has to be stable. ◀

Lemma 4.2. If there is no complementary subspace, $\mathcal{K}_i(A - LC) = (A - LC)\mathcal{K}_i$.

Proof. From (15), let v_{ij}, $j = 1, \cdots, k_i$ span $\mathcal{T}_i$ such that $(A - LC)v_{ij} = \sigma_{ij} v_{ij}$. From Lemma 4.1, let $\mathcal{K}_i v_{kj} = \bar{k}_{ij} v_{ij}$ if $k = i$ or 0 if $k \neq i$. From Assumption 4.1, any element in the state space $\mathcal{X}$ can be represented by $\sum_{i=1}^{q}\sum_{j=1}^{k_i} \alpha_{ij} v_{ij}$. Then,

$$\mathcal{K}_i(A - LC)\Big(\sum_{i=1}^{q}\sum_{j=1}^{k_i} \alpha_{ij} v_{ij}\Big) = \sum_{j=1}^{k_i} \alpha_{ij}\sigma_{ij}\bar{k}_{ij} v_{ij}$$

$$(A - LC)\mathcal{K}_i\Big(\sum_{i=1}^{q}\sum_{j=1}^{k_i} \alpha_{ij} v_{ij}\Big) = \sum_{j=1}^{k_i} \alpha_{ij}\bar{k}_{ij}\sigma_{ij} v_{ij}$$

imply $\mathcal{K}_i(A - LC) = (A - LC)\mathcal{K}_i$. ◀

From Lemma 4.2 and algebraic form of (14),

$$\mathcal{K}_i = -[(A - LC) + (A - LC)^T]^{-1} H_i \tag{17}$$

where $(A - LC) + (A - LC)^T$ is invertible because $A - LC$ has to be stable. By substituting (17) into (13),

$$L^* = \Big(\sum_{i=1}^{q} H_i\Big)^{-1}\Big[\sum_{i=1}^{q} H_i(\gamma P_i + W_i)\Big]C^T V^{-1} \tag{18}$$

Lemmas 4.3 and 4.4 are important projector properties for Lemma 4.5 which will be used to simplify L^* (18).

Lemma 4.3. There exists a state transformation Γ,

$$\begin{bmatrix} \mathcal{T}_1 & \cdots & \mathcal{T}_q \end{bmatrix} = \Gamma \begin{bmatrix} M_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & M_q \end{bmatrix}$$

where $M_i, i = 1, \cdots, q$, are any invertible $k_i \times k_i$ matrices, such that

$$\Gamma^T H_1 \Gamma = \begin{bmatrix} \bar{H}_1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \Gamma^T H_2 \Gamma = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \bar{H}_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\cdots, \quad \Gamma^T H_q \Gamma = \begin{bmatrix} 0 & 0 \\ 0 & \bar{H}_q \end{bmatrix}$$

Proof. Since H_i (6) has a null space $\hat{\mathcal{T}}_i$,

$$\operatorname{Ker} H_1 = \begin{bmatrix} \mathcal{T}_2 & \cdots & \mathcal{T}_q \end{bmatrix} = \Gamma \begin{bmatrix} 0 \\ \hat{M}_1 \end{bmatrix}$$

where $\hat{M}_1$ is a block diagonal matrix with diagonal matrix elements $M_2, \cdots, M_q$, then

$$H_1\Gamma \begin{bmatrix} 0 \\ \hat{M}_1 \end{bmatrix} = 0 \;\Rightarrow\; \Gamma^T H_1 \Gamma \begin{bmatrix} 0 \\ \hat{M}_1 \end{bmatrix} = 0$$

Since $\hat{M}_1$ is not zero by definition and $\Gamma^T H_1 \Gamma$ is symmetric,

$$\Gamma^T H_1 \Gamma = \begin{bmatrix} \bar{H}_1 & 0 \\ 0 & 0 \end{bmatrix}$$

Similarly, $i = 2, \cdots, q$, can be proved. ◀

Lemma 4.4.

$$H_i(\sum_{k=1}^{q} H_k)^{-1} H_j = \begin{cases} H_i & , \quad i = j \\ 0 & , \quad i \neq j \end{cases}$$

Proof. For $i = 1$ and $j = 2$,

$$\begin{aligned}
&\Gamma^T H_1 (\sum_{k=1}^{q} H_k)^{-1} H_2 \Gamma \\
&= (\Gamma^T H_1 \Gamma)(\sum_{k=1}^{q} \Gamma^T H_k \Gamma)^{-1} (\Gamma^T H_2 \Gamma) \\
&= \begin{bmatrix} \bar{H}_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{H}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \bar{H}_q \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 & 0 \\ 0 & \bar{H}_2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & \bar{H}_2 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0
\end{aligned}$$

Therefore, $H_1(\sum_{k=1}^{q} H_k)^{-1} H_2 = 0$ and similarly it can be shown for all cases. ◀

Lemma 4.5. If there is no complementary subspace, $H_i W_i = 0$.

Proof. By multiplying algebraic form of (10) by H_i from left and right, substituting L with (18) and using Lemma 4.4,

$$H_i(\bar{A}_i W_i + W_i \bar{A}_i^T - W_i C^T V^{-1} C W_i) H_i = 0$$

where $\bar{A}_i = A - P_i C^T V^{-1} C$. Since $\bar{A}_i$ is stable [11, 12], the solution is either $W_i = 0$ or $\mathrm{Im} W_i \subseteq \mathrm{Ker} H_i$. Therefore $H_i W_i = 0$. ◀

By using Lemma 4.5, L^* (18) becomes

$$L^* = (\sum_{i=1}^{q} H_i)^{-1} (\sum_{i=1}^{q} \gamma H_i P_i) C^T V^{-1} \tag{19}$$

Theorem 4.6 shows that L^* (19) is consistent with the assumption (15) in the limit and therefore the limiting optimal stochastic multiple-fault detection filter is equivalent to the Beard-Jones fault detection filter.

Theorem 4.6. In the limit, for $i = 1, \cdots, q$,

$$(A - LC)\mathcal{T}_i \subseteq \mathcal{T}_i \tag{20}$$

where $L = (\sum_{i=1}^{q} H_i)^{-1} (\sum_{i=1}^{q} H_i P_i) C^T \bar{V}^{-1}$

Proof. Instead of P_i (9), using its inverse $\Pi_i \triangleq P_i^{-1}$ is a better way to discuss the limiting properties because Π_i has a null space $\hat{\mathcal{T}}_i$ in the limit [11, 12]. Then, the filter gain becomes

$$L = (\sum_{i=1}^{q} H_i)^{-1} (\sum_{i=1}^{q} H_i \Pi_i^{-1}) C^T \bar{V}^{-1} \tag{21}$$

where

$$\begin{aligned}
0 =& \Pi_i A + A^T \Pi_i + \Pi_i (\frac{1}{\gamma} \hat{F}_i \hat{Q}_i \hat{F}_i^T - F_i Q_i F_i^T) \Pi_i \\
& - C^T \bar{V}^{-1} C
\end{aligned} \tag{22}$$

Note that the infinite part of Π_i^{-1} is annihilated by the projector H_i because $\mathrm{Ker}\, H_i = \mathrm{Ker}\, \Pi_i$. For $i = 1$, multiply (20) by Π_2 from the left,

$$\Pi_2 (A - LC) \mathcal{T}_1 = 0$$

because $\mathcal{T}_1$ is in the null space of Π_2 in the limit. By substituting L with (21),

$$\Rightarrow \Pi_2 A \mathcal{T}_1 - \Pi_2 (\sum_{i=1}^{q} H_i)^{-1} (\sum_{i=1}^{q} H_i \Pi_i^{-1}) C^T \bar{V}^{-1} C \mathcal{T}_1 = 0$$

$$\Rightarrow \Pi_2 A \mathcal{T}_1 - \Pi_2 (\sum_{i=1}^{q} H_i)^{-1} H_2 \Pi_2^{-1} C^T \bar{V}^{-1} C \mathcal{T}_1 = 0$$

because $\Pi_2 (\sum_{i=1}^{q} H_i)^{-1} H_{j \neq 2} = 0$ which can be shown similarly to Lemma 4.4.

$$\Rightarrow \Pi_2 A \mathcal{T}_1 - C^T \bar{V}^{-1} C \mathcal{T}_1 = 0$$

which is true by multiplying (22) where $i = 2$ by $\mathcal{T}_1$ from the right. Similarly, it can be shown

$$\Pi_i (A - LC) \mathcal{T}_1 = 0$$

for $i = 3, \cdots, q$. Since $\mathrm{Ker}\, \Pi_2 \cap \cdots \cap \mathrm{Ker}\, \Pi_q = \mathcal{T}_1$ [11, 12],

$$(A - LC) \mathcal{T}_1 \subseteq \mathcal{T}_1$$

and for $i = 2, \cdots, q$, it can be shown similarly. ◀

Remark 4. Lemma 4.5 implies that the projectors H_i, (6), minimize the cost (8). Therefore, (6) are the optimal projectors in the limit. ◀

Remark 5. (19) shows a limiting property of the optimal stochastic multiple-fault detection filter. However, the optimal filter gain can not be derived when

γ is zero because the filter gain depends on the inverse of V which is zero. Therefore, only an approximate Beard-Jones fault detection filter can be derived when γ is small. However, when a full-order Beard-Jones fault detection filter reduces to a few reduced-order filters, these reduced-order filters can be recovered by taking the optimal stochastic single-fault detection filter to the limit [12]. ◀

Remark 6. By combining Lemma 4.5 and $H_i P_i(t_1) H_i = 0$ [11, 12], the optimal stochastic multiple-fault detection filter satisfies a disturbance attenuation problem,

$$\frac{\frac{1}{t_1-t_0}\int_{t_0}^{t_1} \mathrm{tr}\left\{\mathrm{E}\left[\sum_{i=1}^{q} \hat{h}_i \hat{h}_i^T\right]\right\} dt}{\frac{1}{t_1-t_0}\int_{t_0}^{t_1} \mathrm{tr}\left\{\mathrm{E}\left[\sum_{i=1}^{q} h_i h_i^T\right]\right\} dt} = 0$$

in the limit. ◀

5 Example

In this section, a numerical example from [4] shows that the minimization problem produces a fault detection filter when there is a complementary subspace. The system matrices are

$$A=\begin{bmatrix} 0 & 3 & 4 \\ 1 & 2 & 3 \\ 0 & 2 & 5 \end{bmatrix}, C=\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, F_1=\begin{bmatrix} 5 \\ 1 \\ 1 \end{bmatrix}, F_2=\begin{bmatrix} 1 \\ -0.5 \\ 0.5 \end{bmatrix}$$

The power spectral densities are chosen as $Q_1 = Q_2 = 1$ and $V = 10^{-6}\, I$. The disturbance attenuation bound γ is 10^{-6}. The infinite-time minimization problem is solved numerically by using gradient method and the frequency response of the filter shows that the two faults are isolated.

6 Conclusion

The optimal stochastic multiple-fault detection filter is a generalization from the single-fault filter. The residual space of the filter is divided into several subspaces and each subspace is sensitive to only its target fault, but not the nuisance faults, in the sense that the transmission from nuisance faults to the target residual space is small while the transmission from target fault is large. In the limit as the nuisance fault weighting goes to infinity and in the absence of sensor noise and a complementary subspace, this filter is equivalent to a Beard-Jones fault detection filter which puts each fault into an unobservable subspace. This filter has the advantages of the unknown input observer in that it can be designed for robustness and the advantages of the Beard-Jones fault detection filter by being capable of detecting multiple faults in one filter. Although there is additional computation to determine the filter gain, this can be done off-line so that implementation is as straightforward as the Beard-Jones fault detection filter.

References

[1] R. V. Beard, *Failure Accomodation in Linear Systems through Self-Reorganization*, Ph.D. thesis, Massachusetts Institute of Technology, 1971.

[2] H. L. Jones, *Failure Detection in Linear Systems*, Ph.D. thesis, Massachusetts Institute of Technology, 1973.

[3] Mohammad-Ali Massoumnia, "A geometric approach to the synthesis of failure detection filters," *IEEE Transactions on Automatic Control*, vol. AC-31, no. 9, pp. 839–846, Sept. 1986.

[4] John E. White and Jason L. Speyer, "Detection filter design: Spectral theory and algorithms," *IEEE Transactions on Automatic Control*, vol. AC-32, no. 7, pp. 593–603, July 1987.

[5] Randal K. Douglas and Jason L. Speyer, "Robust fault detection filter design," *AIAA Journal of Guidance, Control, and Dynamics*, vol. 19, no. 1, pp. 214–218, Jan–Feb 1996.

[6] Randal K. Douglas and Jason L. Speyer, "$\mathcal{H}_\infty$ bounded fault detection filter," *AIAA Journal of Guidance, Control, and Dynamics*, vol. 22, no. 1, pp. 129–138, Jan–Feb 1999.

[7] Mohammad-Ali Massoumnia, George C. Verghese, and Alan S. Willsky, "Failure detection and identification," *IEEE Transactions on Automatic Control*, vol. AC-34, no. 3, pp. 316–321, Mar. 1989.

[8] Walter H. Chung and Jason L. Speyer, "A game theoretic fault detection filter," *IEEE Transactions on Automatic Control*, vol. AC-43, no. 2, pp. 143–161, Feb. 1998.

[9] George H. Lee, *Least-Squares and Minimax Methods for Filtering, Identification and Detection*, Ph.D. thesis, University of California at Los Angeles, 1994.

[10] Thomas S. Brinsmead, J. S. Gibson, Graham C. Goodwin, George H. Lee, and D. Lewis Mingori, "Fault detection: A quadratic optimization approach," in *Proceedings of the 36th Conference on Decision and Control*, 1997, pp. 3970–3975.

[11] Robert H. Chen and Jason L. Speyer, "Optimal stochastic fault detection filter," in *Proceedings of the American Control Conference*, 1999, pp. 91–96.

[12] Robert H. Chen and Jason L. Speyer, "Residual-sensitive fault detection filter," in *Proceedings of the 7th IEEE Mediterranean Conference on Control and Automation*, 1999.

APPENDIX F

A Decentralized Fault Detection Filter

W.H. Chung and J.L. Speyer

A DECENTRALIZED FAULT DETECTION FILTER

Walter H. Chung and Jason L. Speyer*
Mechanical and Aerospace Engineering Department
University of California, Los Angeles
48-121 Engr. IV
Los Angeles, CA 90095
speyer@seas.ucla.edu or walter.h.chung@aero.org

ABSTRACT

In this paper, we introduce the decentralized fault detection filter, a structure that results from merging decentralized estimation theory with the game theoretic fault detection filter. A decentralized approach may be the ideal way to health monitor large-scale systems, since it decomposes the problem down into (potentially smaller) "local" problems and then blends the "local" results into a "global" result that describes the health of the entire system. The benefits of such an approach include added fault tolerance and easy scalability. An example given at the end of the paper demonstrates the use of this filter for a platoon of cars proposed for advanced vehicle control systems.

Introduction

Observers play a central role in an important class of techniques for fault detection and identification (FDI). Since failures act as nexpected inputs, they will bias the error residuals of any observer designed about the nominal system. Moreover, because of their closed-loop nature, observers are able to maintain nonzero residuals for indefinite periods of time after the occurence of a failure[1], and they possess reduced sensitivity to model mismatch, nonlinearities, and exogenous disturbances inherent to feedback systems.

There are two types of observers currently used for FDI purposes. The first is known as the *Beard-Jones Fault Detection Filter* (White and Speyer, 1987; Massoumnia, 1986; Douglas, 1993). This filter is a variation of the Luenberger Observer in which nonoverlapping invariant subspaces have been built around the reachable subspaces of the failures modelled in the system. The influence of any one of these failures is restricted to its own particular subspace, which allows for simultaneous detection and identification. That is, projecting the error residual onto each of these invariant subspaces one-by-one, a failure is detected when the projection is nonzero and identified by the subspace corresponding the nonzero projection.

The second type of FDI observer is known as the *unknown input observer*. In this observer, the set of modelled faults is divided into two groups: the faults to be detected and the faults that are to be ignored. The former is made distinguishable from the latter

*This work was sponsored by NASA-Ames Research Center Cooperative Agreement No. NCC 2-374, supplement 19, Air Force Office of Scientific Research Grant F49620-97-1-0272, and California Department of Transportation Agreement No. 65H 978, MOU 126

[1] A distinct advantage over open-loop FDI methods (White and Speyer, 1987)

by constructing an output through which the latter set is unobservable. Detection is then achieved when this output is nonzero and identification is trivial because we are only trying to detect one set in the possible presence of the other. The unknown input observer is clearly less capable than the Beard-Jones filter, but its relatively simple structure allows for easy approximation by optimization methods (Ding and Frank, 1989; Chung and Speyer, 1998).

As both of these approaches have become more refined, applications have begun to be seen in the literature for systems as varied as jet engines (Patton and Chen, 1992), missile guidance (Bowman and Speyer, 1987), nuclear reactors (Patton et al., 1991), and automated highways (Douglas et al., 1995, 1996). With the advent of applications, however, new issues related to implementation have come to the forefront. In this paper, we will look at some of the challenges inherent to detecting faults in large-scale systems. For such systems, a *decentralized fault detection filter* may be the logical approach to the problem.

The decentralized fault detection filter is the result of combining the game theoretic fault detection filter of Chung and Speyer (1998) with the decentralized filtering algorithm introduced by Speyer (1979) and extended by Willsky et al. (1982). It approximates the actions of an unknown input observer and is formed by combining the estimates of several "local" estimators (each driven by independent measurement sets). For large-scale systems, it simplifies the health montoring problem by decomposing it down into a collection of smaller problems. For other systems like a platoon of cars (Douglas et al., 1996; Wolfe et al., 1996) or a formation of airplanes, its decentralizedstructure reflects the actual physical structure of the system. And further, it introduces scalability for circumstances such as when a car joins the platoon or when an airplane drops out of formation for repairs. Finally, the decentralize fault detection filter has built in fault tolerance in that sensors can be checked and validated prior to their measurements being blended into the global estimate (Kerr, 1985).

Decentralized Estimation Theory and Its Application to FDI

The General Solution

In this section we will review the basic results of decentralized estimation theory. A detailed examination of this theory is given in Chung and Speyer (1995).

Consider the following system driven by process disturbances w and sensor noise v,

$$\dot{x} = Ax + Bw, \qquad x(0), x \in \mathcal{R}^n, \tag{1}$$

$$y = Cx + v, \qquad y \in \mathcal{R}^m. \tag{2}$$

It is desired to derive an estimate of x. The standard approach is a full-order observer,

$$\dot{\hat{x}} = A\hat{x} + L(y - C\hat{x}), \qquad \hat{x}(0) = 0, \tag{3}$$

which we will refer to as a *centralized estimator*. An alternative to this method is to derive the estimate with a *decentralized estimator*. In the decentralized approach, $\hat{x}$ is found by combining estimates based upon "local" models,

$$\dot{x}^j = A^j x^j + B^j w^j, \qquad x^j \in \mathcal{R}^{n^j}, \qquad (j = 1 \ldots N), \tag{4}$$

$$y^j = E^j x^j + v^j, \qquad y^j \in \mathcal{R}^{m^j}, \qquad (j = 1 \ldots N). \tag{5}$$

Together these local models provide an alternate representation of the original system, which is referred to as the "global" system for purposes of clarification. The vector x is likewise called the "global" state. The number of local systems N is bounded above by the number of measurements in the system, i.e. $N \leq m$.

The global/local decomposition is really of only secondary importance. As Chung and Speyer (1995) argue, there are no real restrictions on how one forms the global and local models. The real key to the decentralized estimation algorithm is the relationship between the global set of measurements y and the N local sets, y^j. The two basic assumptions are that the local sets are simply segments of the global set,

$$y = \begin{Bmatrix} y^1 \\ y^2 \\ \vdots \\ y^N \end{Bmatrix}, \tag{6}$$

and that the local sets can be described in terms of both the local state *and* the global state. In other words, y^j can be given by (5) or by

$$y^j = C^j x + v^j, \qquad (j = 1 \ldots N). \tag{7}$$

Equations 2, 6, and 7 imply that

$$C = \begin{bmatrix} C^1 \\ \vdots \\ C^N \end{bmatrix}$$

and that

$$v = \begin{bmatrix} v^1 \\ \vdots \\ v^N \end{bmatrix}. \tag{8}$$

The decentralized estimation algorithm falls out when we attempt to estimate the global state by first generating estimates of the local systems (4) using the local measurement sets y^j and the local models A^j:

$$\dot{\hat{x}}^j = A^j \hat{x}^j + L^j (y^j - E^j \hat{x}^j), \qquad \hat{x}^j(t_0) = 0, \qquad (j = 1 \ldots N). \tag{9}$$

The global state estimate, $\hat{x}$, is then found via

$$\hat{x} = \sum_{j=1}^{N} \left(G^j \hat{x}^j + h^j \right), \tag{10}$$

where h^j is a measurement-dependent variable propagated by

$$\dot{h}^j = \Phi h^j + (\Phi G^j - \dot{G}^j - G^j \Phi^j) \hat{x}^j, \qquad h^j(0) = 0. \tag{11}$$

The constituent matrices are defined as

$$\begin{aligned} \Phi &:= A - \sum_{j=1}^{N} G^j L^j C^j, \\ \Phi^j &:= A^j - L^j E^j. \end{aligned}$$

The G^j matrices are "blending matrices". Later, we will suggest a method for determining these matrices. In Chung and Speyer (1995), it was found that in order to get the same estimate using either the decentralized or standard centralized algorithms, the local and global gains had to be related via,

$$L = \left[G^1 \ldots G^N \right] \begin{bmatrix} L^1 & 0 & \cdots & 0 \\ 0 & L^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & L^N \end{bmatrix}. \tag{12}$$

In general, however, this condition can not be met because of an insufficient number of equations required to solve for the unknowns.

There is, however, one general class of estimator for which (12) is satified almost automatically. This class is comprised of estimators which take their gains from Riccati solutions, i.e. Kalman Filters (Speyer, 1979; Willsky et al., 1982) or H^{∞} filters (Jang and Speyer, 1994). In this case, the local gains are found from

$$L^j = P^j (E^j)^T (V^j)^{-1}, \tag{13}$$

where, in the case of the Kalman Filter, the matrix, P^j is the solution of the Riccati Equation:

$$\dot{P}^j = A^j P^j + P^j (A^j)^T + B^j W^j (B^j)^T - P^j (E^j)^T (V^j)^{-1} E^j P^j,$$
$$P^j(0) = P_0^j.$$

The matrices, V^j and W^j, are weightings which are taken to be the power spectral densities of the local disturbances, v^j and w^j, which drive the local systems (4,5). For the Kalman Filter it is assumed that v^j and w^j are white, Gaussian signals. The initial condition P_0^j is chosen by the analyst based upon his knowledge of the system. In the global system, the global gain is

$$L = PC^T \mathcal{V}^{-1},$$

where

$$\mathcal{V} = \begin{bmatrix} V^1 & 0 & \cdots & 0 \\ 0 & V^2 & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & V^N \end{bmatrix}, \tag{14}$$

is restricted to a block diagonal form comprised of the local weightings V^j. The matrix P is the solution to the global Riccati Equation,

$$\dot{P} = AP + PA^T + B\mathcal{W}B^T - PC^T \mathcal{V} CP, \qquad P(0) = P_0.$$

The blending matrix solution is then,

$$G^j = P(S^j)^T (P^j)^{-1} \qquad j = 1, \ldots, N, \tag{15}$$

where S^j is any matrix such that

$$C^j = E^j S^j. \tag{16}$$

One can, in fact, always take $S^j = (E^j)^\dagger C^j$ where $(E^j)^\dagger$ is the pseudo-inverse of E^j (Willsky et al., 1982). Note that the solutions for G^j will always exist for Riccati-based observers so long as P^j is invertible or, equivalently, positive-definite. This will always be the case if the triples, (C^j, A^j, B^j), are controllable and observable for each of the local systems.

Implications for Detection Filters

The analysis of the previous section implies that we will be able to form a decentralized fault detection filter in the general case only if we are able to find a Riccati-based observer which is equivalent to a Beard-Jones Filter or unknown input observer. The most direct way to achieve this is to find a linear-quadratic optimization problem which is equivalent to the fault detection and identification problem. This is an analog of the famous inverse optimal control problem first posed by Kalman (1964). In Chung and Speyer (1998), however, it is shown that the Beard-Jones Filter gains do not correspond with those derived from linear-quadratic problems. An indirect way to get a Riccati-based observer is to pose a linear-quadratice optimization problem which closely mimics the fault detection problem. Such a problem was posed and solved in Chung and Speyer (1998), and we will review the solution found there in the next section.

The Approximate Fault Detection and Identification Problem

Problem Formulation

Consider the system given by (1,2) with the further assumption that the state matrices have sufficient smoothness to guarantee the existence of derivatives various order. Beard (1971) showed that failures in the sensors and actuators, and unexpected changes in the plant dynamics can be modeled as additive signals,

$$\dot{x} = Ax + Bw + F_1\mu_1 + \cdots + F_q\mu_q. \tag{17}$$

Let n be the dimension of the state-space. The $n \times p_i$ matrix F_i, $i = 1 \cdots q$, is called a *failure map* and represents the directional characteristics of the ith fault. The $p_i \times 1$ vector μ_i is the *failure signal* and represents the time dependence of the failure. It will always be assumed that each F_i is monic, i.e. $F_i\mu_i \neq 0$ for $\mu_i \neq 0$. See (Douglas, 1993; Chung and Speyer, 1998) for further details on how to model failures. Throughout this paper, we will refer to μ_1 as the "target fault" and the other faults μ_j, $j = 2 \cdots q$, as the "nuisance faults".

Without loss of generality, we can represent the entire set of nuisance faults (and, if desired, the disturbance w) with a single map F_2 and vector μ_2:

$$\dot{x} = Ax + F_1\mu_1 + F_2\mu_2.$$

Suppose that it is desired to detect the occurrence of the failure, μ_1, in spite of the measurement noise, v, and the possible presence of the nuisance faults, μ_2. The Beard-Jones Filter solves this problem by picking the gain to a standard Luenberger Observer,

$$\dot{\hat{x}} = A\hat{x} + L(y - C\hat{x}), \tag{18}$$

so that the reachable subspaces of μ_1 and μ_2 are in separate and nonintersecting invariant subspaces. Thus, with a properly chosen projector H we can project the filter residual, $(y - C\hat{x})$, onto the orthogonal complement of the invariant subspace containing μ_2 and get a signal,

$$z = H(y - C\hat{x}), \tag{19}$$

such that

$$z = 0 \qquad \text{when } \mu_1 = 0 \text{ and } \mu_2 \text{ is arbitrary.} \tag{20}$$

To be useful for FDI, z must also be such that

$$z \neq 0 \qquad \text{when } \mu_1 \neq 0. \tag{21}$$

If we restrict ourselves to time-invariant systems, (21) will be equivalent to requiring the transfer function matrix between $\mu_1(s)$ and $z(s)$[2] to be *left-invertible*. Left-invertibility, however, is a severe restriction, and it has no analog for the general time-varying systems that we want to consider here. Previous researchers (Douglas, 1993; Massoumnia et al., 1989) have, in fact, only required that the mapping from $\mu_1(t)$ to $z(t)$ be *input observable*, i.e. $z \neq 0$ for any μ_1 that is a step input. It is then argued (Massoumnia et al., 1989) that with input observability z will be nonzero for "almost any" μ_1, since μ_1 is unlikely to remain in the kernal of the mapping to z for all time.

[2] $\mu_1(s)$ and $z(s)$ are the LaPlace Transforms of the time-domain signals $\mu_1(t)$ and $z(t)$.

We formulate the approximate detection filter design problem by requiring input observability and relaxing the requirement for strict blocking that is implied by (20). We, instead, only require that the transmission of the nuisance fault be bounded above by a pre-set level, $\gamma > 0$:

$$\frac{\|z\|^2}{\|\mu_2\|^2} \leq \gamma. \tag{22}$$

Equation 22 is identical to the disturbance attenuation problem from robust control theory. We refer to the solution to the approximate detection filter problem as the *game theoretic fault detection filter*.

We complete our formulation of the disturbance attenuation problem for fault detection by constructing the projector H that determines the failure signal z. For time-invariant systems, this projector is constructed to map the reachable subspace of μ_2 to zero (Beard, 1971; Douglas, 1993), i.e.

$$H = I - C\hat{F}\left[(C\hat{F})^T C\hat{F}\right]^{-1}(C\hat{F})^T, \tag{23}$$

where

$$\hat{F} = \left[A^{\beta_1} f_1, \ldots, A^{\beta_{p_2}} f_{p_2}\right]. \tag{24}$$

The vector f_i, $i = 1 \cdots p_2$, is the ith column of F_2, and the integer β_i is the smallest natural number such that $CA^{\beta_i} f_i \neq 0$. The time-varying extension of this result is

$$H = I - C\hat{F}(t)\left[(C\hat{F}(t))^T C\hat{F}(t)\right]^{-1}(C\hat{F}(t))^T. \tag{25}$$

The columns of the matrix,

$$\hat{F}(t) = \left[b_1^{\beta_1}(t), \ldots, b_{p_2}^{\beta_{p_2}}(t)\right], \tag{26}$$

are constructed with the Goh Transformation (Chung and Speyer, 1998):

$$b_i^1(t) = f_i(t), \tag{27}$$

$$b_i^j(t) = A(t) b_i^{j-1}(t) - \dot{b}_i^{j-1}. \tag{28}$$

In the time-varying case, β_i is the smallest integer for which the iteration above leads to a vector, $b_i^{\beta_k}(t)$, such that $C(t)b_i^{\beta_k}(t) \neq 0$ for all $t \in [t_0, t_1]$. It will be assumed that $A(t), C(t)$, and $F_2(t)$ are such that β_i exists. Since the state-space has dimension n, β_i is such that $0 \leq \beta_i \leq n-1$.

We are now ready to discuss the conditions under which the solution to (22) will also generate an input observable mapping from μ_1 to z. The key requirement is that the system be *output separable*. That is, F_1 and F_2 must be linearly independent and remain so when mapped to the output space by C and A. For time-invariant systems, the test for output separability is

$$\operatorname{rank}\left[CA^{\delta_1}\bar{f}_1, \ldots, CA^{\delta_{p_1}}\bar{f}_{p_1}, CA^{\beta_1}f_1, \ldots, CA^{\beta_{p_2}}f_{p_2}\right] = p_1 + p_2. \tag{29}$$

As in (24), f_i is the ith column of F_2, and β_i is the the smallest integer such that $CA^{\beta_i}f_i \neq 0$. Similarly, $\bar{f}_j$ is the jth column of F_1, and δ_j is the smallest integer such that $A^{\delta_j}\bar{f}_j \neq 0$. The integer sum, $p_1 + p_2$, is the total number of columns in F_1 and F_2.

For time-varying systems, the output separability test becomes

$$\operatorname{rank}\left[C(t)\bar{b}_1^{\delta_1}(t), \ldots, C(t)\bar{b}_{p_1}^{\delta_{p_1}}(t), C(t)b_1^{\beta_1}(t), \ldots, C(t)b_{p_2}^{\beta_{p_2}}(t)\right] = p_1 + p_2, \quad \forall t \in [t_0, t_1], \tag{30}$$

where the vectors, $b_i^{\beta_i}$ and $\bar{b}_j^{\delta_j}$, are found from the iteration defined by (27) and (28). The initial vector, $\bar{b}_j^1$, is set equal to the jth column of F_1, and b_i^1 is initialized as the ith column of F_2.

The following proposition given in Chung and Speyer (1998), connects output separability to input observability and shows the importance of the monicity assumption:

Theorem 1 *Suppose that a given filter satisfies (22) and generates the failure signal, z, given by (19). If F_1 and F_2 are output separable and F_1 is monic, then the mapping, $\mu_1(t) \mapsto z(t)$, is input observable.*

A Game Theoretic Solution

We now turn our attention to the disturbance attenuation problem implied by (22). We begin by defining a disturbance attenuation function (Rhee and Speyer, 1991),

$$D_{af} = \frac{\int_{t_0}^{t_1} \|HC(x-\hat{x})\|_Q^2 dt}{\int_{t_1}^{t_2}\left[\|\mu_2\|_{M^{-1}}^2 + \|v\|_{V^{-1}}^2\right]dt + \|x(t_0) - \hat{x}_0\|_{P_0}^2}. \tag{31}$$

D_{af} is simply a ratio of the outputs over the disturbances . Equation 31 is patterned roughly after (22). We have added the sensor noise, v, and the initial error, $x(t_0) - \hat{x}_0$, to the set of disturbance signals to inject tradeoffs for noise rejection and settling time into the problem. M, V, Q, and P_0 are weighting matrices. Note that we do not include the target fault μ_1 at this stage of the design problem, since we are now focusing on nuisance blocking. Our only concern with μ_1 is that it be visible at the output, which is what Proposition 1 guarantees. The disturbance attenuation problem is to find the estimate $\hat{x}$ so that for all μ_2, $v \in L_2[t_1, t_2]$, and $x(t_0) \in \mathcal{R}^n$,

$$D_{af} \leq \gamma.$$

The positive real number γ is called the *disturbance attenuation bound.* (C, A) will always be assumed to be an observable pair.

To solve this problem, we convert (31) into a cost function,

$$J = \int_{t_0}^{t_1} \left[\|HC(x - \hat{x})\|_Q^2 - \gamma \left(\|\mu_2\|_{M^{-1}}^2 + \|y - Cx\|_{V^{-1}}^2 \right) \right] dt - \|x(t_0) - \hat{x}_0\|_{\Pi_0}^2, \tag{32}$$

where we have used (2) to rewrite the measurement noise term. Note that we have also rewritten the initial error weighting, defining $\Pi_0 := \gamma^{-1} P_0$. The disturbance attenuation problem is then solved via the differential game,

$$\min_{\hat{x}} \max_{y} \max_{\mu_2} \max_{x(t_0)} J \leq 0, \tag{33}$$

subject to

$$\dot{x} = Ax + F_2 \mu_2, \tag{34}$$

$$y = Cx + v.$$

The solution to this problem (Chung and Speyer, 1998) turns out to be a Luenberger Observer,

$$\dot{\hat{x}} = A\hat{x} + \gamma \Pi^{-1} C^T V^{-1} (y - C\hat{x}), \qquad \hat{x}(t_0) = \hat{x}_0, \tag{35}$$

whose gain is taken from the solution to a Riccati Equation,

$$-\dot{\Pi} = A^T \Pi + \Pi A + \frac{1}{\gamma} \Pi F_2 M F_2^T \Pi + C^T (HQH - \gamma V^{-1}) C \qquad \Pi(t_0) = \Pi_0. \tag{36}$$

In many cases, it is desired to extend finite-time solutions of game theoretic problems to the steady-state condition. Whenever it is possible to find such a solution, the optimal estimator will be given by (35) with Π being the solution of the algebraic Riccati Equation,

$$0 = A^T\Pi + \Pi A + \frac{1}{\gamma}\Pi F_2 M F_2^T \Pi + C^T(HQH - \gamma V^{-1})C. \tag{37}$$

However, unlike linear quadratic optimal control problems, there are no conditions which guarantee the existence of a unique, nonnegative definite, stabilizing solution to the steady-state Riccati Equation, except in the special case where A is asymptotically stable (Green and Limebeer, 1995).

The Decentralized Fault Detection Filter

Given the results of the previous two sections, we now propose a decentralized fault detection filtering algorithm. The essential idea is to implement the Riccati-based game theoretic fault detection filter as a decentralized estimator. An overview of the procedure is as follows:

1. Identify the sensors and actuators which must be monitored at the global level, i.e. define the target faults for the global filter.
2. Identify the faults which should be included in the global nuisance set. The remaining faults should be monitored at the local levels.
3. Derive global and local models for the system including failure maps. Chung and Speyer (1998) contains a brief discussion about this process. We will demonstrate one method in which the local models are derived from the global model via a minimum realization.
4. Design game theoretic fault detection filters for the local and global systems. Solve the corresponding Riccati equations and store the solutions for later use.
5. Determine the blending solutions G^j from Equation 15.
6. Propagate the local estimates $\hat{x}^j$ and vectors h^j and then use the decentralized estimation algorithm (10) to derive a global estimate, $\hat{x}$.
7. Determine the global failure signal from $(y - C\hat{x})$ where y is the total measurement set, C is the global measurement matrix, and $\hat{x}$ is the global fault detection filter estimate just derived.

We will now apply these steps in an example.

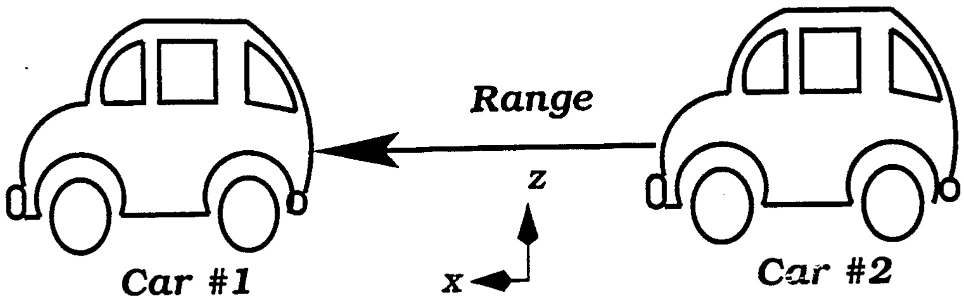

Figure 1. Two-Car Platoon with Range Sensor

Range Sensor Fault Detection in a Platoon of Cars

Problem Statement

We will now examine the utility of the decentralized approach to FDI by working through an example. The problem that we will look at involves the detection of failures within a system of two cars traveling as a platoon (See Figure 1). The cars are controlled to maintain a uniform speed and constant separation. The platoon is the central component of automated highway schemes in which groups of cars line up single file and travel as a unit. The objective is to eliminate the backup caused by the interaction of individual vehicles maneuvering across highway lanes (Douglas et al., 1995, 1996). The viability of the platooning scheme, however, will depend on many factors, not the least of which are reliability and safety.

The FDI schemes that we have examined to this point are capable of monitoring individual cars, but may not be ideal for monitoring elements that deal with the interactions between cars. For example, to maintain uniform speed throughout the platoon and to keep the spacing between the cars constant, additional sensors will be needed to measure the relative speed and the relative distance, or "range", between the cars. In order to detect a failure in the range sensor using analytical redundancy, however, it is necessary to have a dynamical relationship between the range sensor and other sensors on the vehicles. Range, however, involves the dynamics of both of the cars and so would require a higher-order model for its detection filter.

While this is not necessarily prohibitive, it does not make use of the many different state estimates that are already being propagated throughout the platoon. The sensors on each of the cars, for instance, will be monitored by detection filters, and it is more than likely that a state estimate would also be generated by the vehicles' control loops. Given these pre-existing estimates, it seems logical to make use of the decentralized estimation algorithm to carry out range sensor fault detection.

System Dynamics and Failure Modeling

Our example starts with the car model used in Douglas et al. (1995). In this model, the nonlinear, six degree-of-freedom dynamics of an representative automobile are linearized about a straight, level path at a speed of 25 meters/sec (roughly 56 miles per hour). The linearized equations are found to decouple nicely into latitudinal and longitudinal dynamics, much like an airplane. Moreover, the linearized equations can be further reduced by eliminating "fast modes" and actuator states. For simplicity, we will only use the longitudinal dynamics which we represent as

$$\dot{x} = A^L x,$$

$$y = C^L x,$$

where the superscript "L" stands for "longitudinal." The vehicle states are

$$x = \left\{ \begin{array}{c} m_a \\ \omega_e \\ v_x \\ v_z \\ z \\ q \\ \theta \end{array} \right\} \begin{array}{l} \textit{engine air mass } (\mathrm{kg}) \\ \textit{engine speed } (\mathrm{rad/sec}) \\ \textit{long. velocity } (\mathrm{m/sec}) \\ \textit{vertical velocity } (\mathrm{m/sec}) \\ \textit{vertical position } (\mathrm{m}) \\ \textit{pitch rate } (\mathrm{rad/sec}) \\ \textit{pitch } (\mathrm{rad}) \end{array} \tag{38}$$

and are propagated by the state matrix,

$$A^L = \begin{bmatrix} -0.087694 & 0.0038094 & -0.12133 & -0.010701 & 3.9941 & 42.617 & 1.2879 \\ 0.032194 & -1.6765 & 57.123 & 7.2346 & 26.27 & -665.78 & 496.6 \\ 4.6169e-05 & -0.021736 & -22.56 & 0.11478 & -0.00095051 & 7.7651e-05 & -4.5754e-05 \\ -0.075512 & 7.7689 & -301.66 & -38.647 & -137.16 & 3612 & -2816.7 \\ -0.096212 & -0.073026 & 2.498 & 0.2312 & 0.89067 & -19.054 & 9.0737 \\ -0.94943 & -0.26102 & -0.20407 & -0.067025 & -0.41229 & -2.4689 & 0.16425 \\ -0.27186 & 0.92418 & 0.12024 & 0.19024 & -0.010912 & -1.302 & -1.434 \end{bmatrix}. \tag{39}$$

The measurements are

$$\underline{y} = \left\{ \begin{array}{c} m_a \\ \omega_e \\ \dot{v}_x \\ \dot{v}_z \\ q \\ \overline{\omega}_f \\ \overline{\omega}_r \end{array} \right\} \begin{array}{l} \textit{engine air mass } (\mathrm{kg}) \\ \textit{engine speed } (\mathrm{rad/sed}) \\ \textit{long. acceleration } (\mathrm{m/sec^2}) \\ \textit{heave acceleration } (\mathrm{m/sec^2}) \\ \textit{pitch rate } (\mathrm{rad/sec}) \\ \textit{front symmetric wheel speed } (\mathrm{rad/sec}) \\ \textit{rear symmetric wheel speed } (\mathrm{rad/sec}) \end{array} \tag{40}$$

with the corresponding measurement matrix,

$$
C^L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0713 & -0.8177 & 0.5934 & 6.7786 & 16.8068 & 1.5162 \\ 0 & -0.0020 & 0.0221 & -3.5646 & -40.4210 & -9.0765 & -0.8141 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 7.1220 & -4.5806 & -51.9152 & 58.8718 & 5.1944 \\ 0 & 0.0888 & 5.9738 & -3.5782 & -40.5542 & -56.4109 & -4.9773 \end{bmatrix}. \tag{41}
$$

The rear and front symmetric wheel speeds are states that were eliminated when the fast modes were factored out of the linearized system.

In order to build a detection filter for the range sensor, we need to use (38-41) to build state space models for the platoon,

$$
\begin{aligned}
\dot{\eta} &= A\eta + F_1\mu_1 + F_2\mu_2, \\
y &= C\eta,
\end{aligned}
$$

and the two individual cars,

$$
\begin{aligned}
\dot{\eta}^1 &= A^1\eta^1 + F_1^1\mu_1^1 + F_2^1\mu_2^1, \\
y^1 &= E^1\eta^1, \\
\dot{\eta}^2 &= A^2\eta^2 + F_1^2\mu_1^2 + F_2^2\mu_2^2, \\
y^2 &= E^2\eta^2.
\end{aligned}
$$

We will build up our models with the following steps:

1. Using (38-41), we will derive the global state matrices, A and C.
2. Using the modelling techniques described in Douglas (1993) and Chung and Speyer (1998), we will determine the failure maps, F_i.
3. We will then obtain the local state matrices, A^i, E^i, and F_j^i, from the minimum realization of the triples (C^1, A, F_2) and (C^2, A, F_2).

Our general strategy is to derive the global equation first and then get the local equations from decompositions based upon observability and controllability. While this is by no means the only way to obtain the global and local representations of a system, it is a logical method that can be applied to any problem.

The obvious way the get the global matrices, A and C, is to form block diagonal composite matrices with A^L and C^L repeated on the diagonal, i.e.

$$A' = \begin{bmatrix} A^L & 0 \\ 0 & A^L \end{bmatrix}, \qquad C' = \begin{bmatrix} C^L & 0 \\ 0 & C^L \end{bmatrix}.$$

This, however, is not sufficient, since there is no way to describe the range, R, between the two vehicles with the given states, (38). Range is the relative distance between the cars,

$$R = x^1 - x^2,$$

where x^i is the longitudinal displacement of car i. Displacement, however, is not a state of the vehicle (38). We must, therefore, add a range state to the platoon dynamics, using the equation,

$$\dot{R} = v_x^1 - v_x^2.$$

The end result is that the platoon will be a fifteen-state system,

$$\eta = \left\{ \begin{array}{c|l} m_a^1 & \textit{engine air mass}\ \text{(kg) - Car\#1} \\ \omega_e^1 & \textit{engine speed}\ \text{(rad/sec) - Car\#1} \\ v_x^1 & \textit{long. velocity}\ \text{(m/sec) - Car\#1} \\ v_z^1 & \textit{vertical velocity}\ \text{(m/sec) - Car\#1} \\ z^1 & \textit{vertical position}\ \text{(m) - Car\#1} \\ q^1 & \textit{pitch rate}\ \text{(rad/sec) - Car\#1} \\ \theta^1 & \textit{pitch}\ \text{(rad) - Car\#1} \\ m_a^2 & \textit{engine air mass}\ \text{(kg) - Car\#2} \\ \omega_e^2 & \textit{engine speed}\ \text{(rad/sec) - Car\#2} \\ v_x^2 & \textit{long. velocity}\ \text{(m/sec) - Car\#2} \\ v_z^2 & \textit{vertical velocity}\ \text{(m/sec) - Car\#2} \\ z^2 & \textit{vertical position}\ \text{(m) - Car\#2} \\ q^2 & \textit{pitch rate}\ \text{(rad/sec) - Car\#2} \\ \theta^2 & \textit{pitch}\ \text{(rad) - Car\#2} \\ R & \textit{Range}\ \text{(m).} \end{array} \right\}$$

The corresponding state matrix is

$$A = \begin{bmatrix} A^L & 0 \\ 0 & A^L \\ E_1 & -E_1 \end{bmatrix}, \tag{42}$$

$$E_1 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The measurement matrix is

$$C = \begin{bmatrix} C^L & 0 \\ 0 & \begin{matrix} C^L & 0 \\ 0 & 1 \end{matrix} \end{bmatrix} = \begin{bmatrix} C^1 \\ C^2 \end{bmatrix}, \tag{43}$$

where C^1 and C^2 can be inferred from (43). Finally, the local measurement sets are

$$y^1 = \begin{Bmatrix} m_a^1 \\ \omega_e^1 \\ \dot{v}_x^1 \\ \dot{v}_z^1 \\ q^1 \\ \overline{\omega}_f^1 \\ \overline{\omega}_r^1 \end{Bmatrix} \begin{array}{l} \textit{engine air mass } (\text{kg})\text{- Car\#1} \\ \textit{engine speed } (\text{rad/sed})\text{- Car\#1} \\ \textit{long. acceleration } (\text{m/sec}^2)\text{- Car\#1} \\ \textit{heave acceleration } (\text{m/sec}^2)\text{- Car\#1} \\ \textit{pitch rate } (\text{rad/sec})\text{- Car\#1} \\ \textit{front symmetric wheel speed } (\text{rad/sec}) \text{ - Car\#1} \\ \textit{rear symmetric wheel speed } (\text{rad/sec}) \text{ - Car\#1.} \end{array}$$

and

$$y^2 = \begin{Bmatrix} m_a^2 \\ \omega_e^2 \\ \dot{v}_x^2 \\ \dot{v}_z^2 \\ q^2 \\ \overline{\omega}_f^2 \\ \overline{\omega}_r^2 \\ R \end{Bmatrix} \begin{array}{l} \textit{engine air mass } (\text{kg})\text{- Car\#2} \\ \textit{engine speed } (\text{rad/sed})\text{- Car\#2} \\ \textit{long. acceleration } (\text{m/sec}^2)\text{- Car\#2} \\ \textit{heave acceleration } (\text{m/sec}^2)\text{- Car\#2} \\ \textit{pitch rate } (\text{rad/sec})\text{- Car\#2} \\ \textit{front symmetric wheel speed } (\text{rad/sec}) \text{ - Car\#2} \\ \textit{rear symmetric wheel speed } (\text{rad/sec}) \text{ - Car\#2} \\ \textit{range } (\text{rad/sec}). \end{array}$$

Our ultimate objective is to design a filter which will detect a range sensor fault in the presence of potential failures in the other sensors. In an actual health monitoring system, we would design the global filter to block out all of the nuisance faults that are output separable from the range sensor fault and then rely upon the local filters to monitor the remaining faults. Given the size of our example, however, the full analysis required to do a detailed design would clutter our presentation. We will, therefore, limit ourselves to constructing only one local filter and will choose simple nuisance sets at both the global and local levels.

For this example, we choose to monitor the front symmetric wheel speed sensor at the local level. The nuisance set is then chosen to be the engine air mass sensor and the heave accelerometer. At the global level, the range sensor has already been designated as the target fault. We, therefore, complete the problem definition by choosing the engine speed sensor and longitudinal accelerometer as the global nuisance set. There is no particular significance attached to any of our choices for the nuisance and target sets, aside from the choice of the range sensor as the global target fault.

Following standard modelling techniques Douglas (1993); Chung and Speyer (1998), we construct the two engine speed sensor failure maps $F_{\omega_e^1}$ and $F_{\omega_e^2}$. To save space we do not list these matrices out explicitly. The interested reader can refer to (Chung, 1997).

To complete the problem we also need to construct maps for the accelerometer failures, $F_{\dot{v}_x^1}$ and $F_{\dot{v}_x^2}$, and the range sensor, F_R. For the local filters, failure maps need to be constructed for the airmass sensors, $F_{m_a^1}$ and $F_{m_a^2}$, heave accelerometers, $F_{\dot{v}_z^1}$ and $F_{\dot{v}_z^2}$, and front wheel speed sensors, $F_{\overline{\omega}_f^1}$ and $F_{\overline{\omega}_f^2}$. A quick application of (29) will show that all of our failure sets are output separable.

We are now in position to generate the local state equations. The local dynamics for car #1 come from the minimum realization of $(C^1, A, \left[F_{m_a^1} \; F_{\dot{v}_z^1} \right])$. The corresponding matrices are

$$A^1 = \begin{bmatrix} -0.087694 & 0.0038094 & -0.12133 & -0.010701 & 3.9941 & 42.617 & 1.2879 \\ 0.032194 & 1.6765 & 57.123 & 7.2346 & 26.27 & -665.78 & 496.6 \\ 4.6169e-05 & -0.021736 & -22.56 & 0.11478 & -0.00095051 & 7.7651e-05 & -4.5754e-05 \\ -0.075512 & 7.7689 & -301.66 & -38.647 & -137.16 & 3612 & -2816.7 \\ -0.096212 & -0.073026 & 2.498 & 0.2312 & 0.89067 & -19.054 & 9.0737 \\ -0.94943 & -0.26102 & -0.20407 & -0.067025 & -0.41229 & -2.4689 & 0.16425 \\ -0.27186 & 0.92418 & 0.12024 & 0.19024 & -0.010912 & -1.302 & -1.434 \end{bmatrix},$$

$$E^1 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -0.00039519 & 0.18605 & 0 & -0.98251 & 0.008136 & -0.00066466 & 0.00039164 \\ 0.0043561 & -0.014182 & 0 & -0.090334 & -0.2118 & 11.266 & -14.31 \\ 0.00015951 & -0.00067636 & 0 & -0.0048006 & -4.0642 & -41.318 & -2.4264 \\ -0.00014266 & -0.97872 & 0 & -0.18537 & 0.0016064 & 0.024547 & -0.084511 \\ -0.00030256 & 0.0016942 & 0 & 0.0069288 & 1.4478 & -34.102 & -71.377 \\ 0.0009564 & -0.0038718 & 0 & -0.019192 & 2.1041 & -55.207 & 42.987 \end{bmatrix},$$

$$F_{m_a^1}^1 = \begin{bmatrix} 0 & -0.12133 \\ 0 & 57.1230 \\ 1 & -22.5605 \\ 0 & -301.6586 \\ 0 & 2.4980 \\ 0 & -0.2041 \\ 0 & 0.12024 \end{bmatrix}, \qquad F_{\dot{v}_z^1}^1 = \begin{bmatrix} 7.9031 & -1.6879 \\ -0.0007 & -0.0213 \\ 0 & 0 \\ -0.0048 & -0.0057 \\ -0.1760 & -0.7911 \\ -0.0068 & -7.4136 \\ -0.0003 & -2.1388 \end{bmatrix}.$$

The model for Car #2 is similarly found by obtaining the minimum realization of $(C^2, A, \left[F_{m_a^2} \; F_{\dot{v}_z^2} \right])$. The corresponding matrices are

$$A^2 = \begin{bmatrix} -0.26387 & -0.27372 & 0.97419 & -0.040683 & 0 & 0 & 0 & 0 \\ 0.28256 & 0.2607 & 0.042752 & 1.0237 & 0 & 0 & 0 & 0 \\ -12.546 & -12.054 & -1.4539 & -0.79488 & -0.0025104 & 0.00016431 & 0.00013564 & 0.034164 \\ -28.279 & -27.514 & -2.1059 & -3.0468 & 0.0048048 & 8.1292 & 6.7111 & -0.065389 \\ 195.07 & 193.92 & -2.3745 & 38.898 & -0.19848 & -152.87 & -126.21 & 2.7044 \\ 3.8593 & 4.598 & 0.3571 & 0.51933 & -4.1413e-06 & -21.332 & -18.419 & 5.6359e-05 \\ -4.0915 & -4.8456 & -0.37617 & -0.54711 & 3.0926e-05 & 22.827 & 17.824 & -0.00042087 \\ -2654.8 & -2639.1 & 32.315 & -529.37 & -304.44 & 2080.5 & 1717.5 & -57.774 \end{bmatrix},$$

$$E^2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.99731 & 0 & 0 & 0.073283 \\ 0 & 0 & 0 & 0 & 0.073283 & 0 & 0 & -0.99731 \\ -12.008 & -11.76 & -0.54089 & -1.6668 & 0.0052249 & 5.2402 & 4.3261 & -0.071106 \\ 5.9034 & 6.9535 & 0.53293 & 0.79148 & -0.00014384 & -31.362 & -25.727 & 0.0019575 \\ -0.011223 & 0.011321 & -0.73157 & -0.68157 & 0 & -0.0019527 & -0.0016121 & 0 \\ -43.291 & -39.922 & -8.6999 & 1.9601 & 0 & -40.162 & -33.156 & 0 \\ 40.011 & 39.775 & -0.48704 & 7.9783 & 0.0065064 & -31.356 & -25.886 & -0.088546 \\ 0.69369 & -0.71973 & -0.014645 & -0.0075736 & 0 & -0.017553 & -0.014491 & 0 \end{bmatrix},$$

$$F^2_{m^2_a} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0.9973 & 0.0002 \\ 0 & 0 \\ 0 & 0 \\ 0.0733 & -307.8575 \end{bmatrix}, \qquad F^2_{\dot{v}^2_z} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -5.0327 & -4.9282 \\ 6.0961 & -6.2254 \\ 0 & 0 \end{bmatrix}.$$

With all of these system matrices in place, we can now form the residual projectors H needed generate the failure signal, z. In the global filter, we define

$$\hat{F} = \left[F_{\omega^1_t} \; F_{\dot{v}^1_z} \; F_{\omega^2_t} \; F_{\dot{v}^2_z} \right].$$

In the local filters, we define

$$\hat{F}^i = \left[F^i_{m^i_a} \; F^i_{\dot{v}^i_z} \right] \qquad i = 1, 2$$

The projectors H and H^i are then found by applying (23). Again, we do not show either of these matrices explicitly to save space.

Decentralized Fault Detection Filter Design

We will first design filters for the local systems. As with all Riccati-based filters, the central step in the process is in obtaining a solution to the appropriate Riccati Equation. For simplicity, we will use the steady-state version. Typically, one iterates on the design by trying various combinations of weightings until a Riccati solution is found which leads to a filter that gives the best tradeoff between target fault transmission and nuisance fault attenuation. For this example, it was found that

$$M^1 = 10 \times I_7, \qquad V^1 = \mathrm{diag}\left[1\ 1\ 10\ 1\ 1\ 1\ 1\right],$$

$$Q^1 = I_7, \qquad \gamma = 0.18$$

leads to the filter for Car #1 depicted in Figure 2. The minimum separation over frequency is only 35 dB, but the filter has particularly good separation in the low frequency range. For Car #2, the same weightings, adjusted for the different dimensions of the Car #2 dynamics,

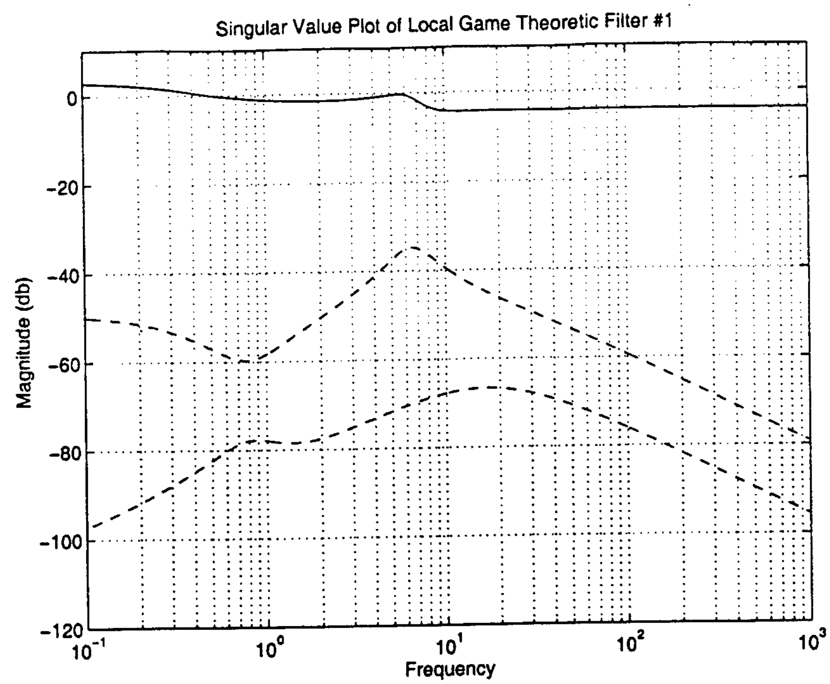

Figure 2. Platoon Example - Signal Transmission in the Local Detection Filter on Car # 1 (accelerometer fault transmission shown with solid line, nuisance fault transmission shown with dashed line)

$$M^2 = 10 \times I_8, \qquad V^2 = \text{diag}\left[1\ 1\ 10\ 1\ 1\ 1\ 1\ 1\right],$$

$$Q^2 = I_8, \qquad \gamma = 0.18,$$

leaa tò a filter with the performance depicted in Figure 3. Finally, for the global system, a fault detection filter for range sensor health monitoring in the platoon is found by solving the corresponding Riccati Equation with the weightings:

$$\gamma V^{-1} = I_{17}, \qquad Q = I_{17},$$

$$M = 100 \times I_8, \qquad \gamma = 0.18.$$

The resulting filter has the properties depicted in Figure 4. The decentralized implementation that we proposed in the previous section should also exhibit this level of performance. As a check, a simple time domain simulation was run comparing the response of the residual signal when the system is driven by the target fault (a step failure of the range sensor) to when it is driven by a nuisance fault (a step failure of the longitudinal accelerometer on Car #1). Because we are using Riccati-based estimators, the blending matrices, G^J,

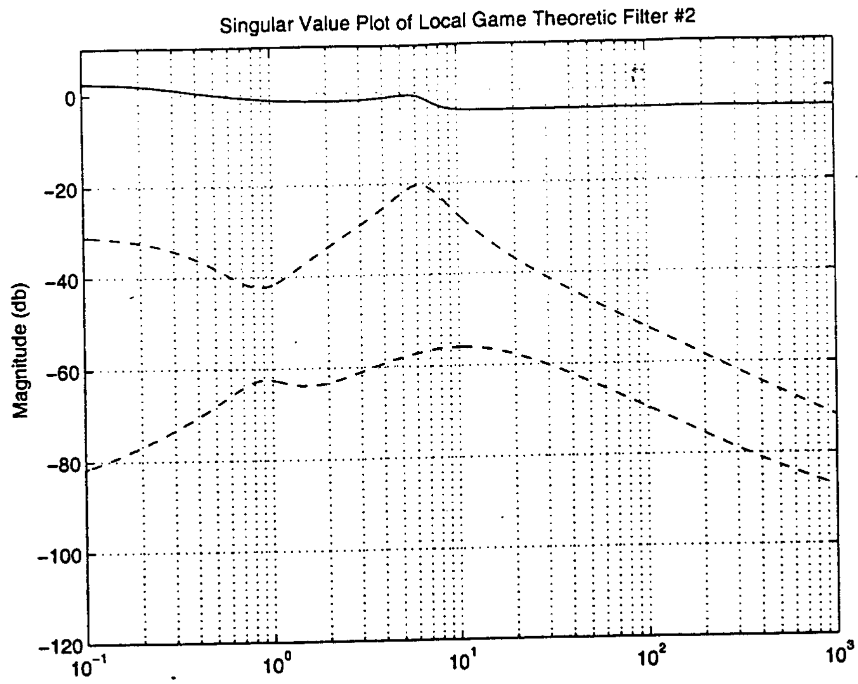

Figure 3. Platoon Example - Signal Transmission in the Local Detection Filter on Car # 2 (accelerometer fault transmission shown with solid line, nuisance fault transmission shown with dashed line)

are given by (15). The connecting matrices S^j are taken to be the pseudo-inverses of E^j. As Figure 5 shows, the resulting decentralized fault detection filter does a good job of distinguishing the target fault from the nuisance fault.

Remark 2 *It must be noted that we have assumed that the lead car will transmit its measurements* y^1, *its local state estimates* $\hat{\eta}^2$, *and the vector* h^1 *back to car #2 so that the latter can form the global estimate via the decentralized estimation algorithm. Transmission issues and limitations, quite obviously, open up the potential for new problems. We have also assumed that each car will have stored on-board the needed Riccati solutions for all likely scenarios.*

Conclusions

In this paper, we have introduced a decentralized fault detection filter which provides an alternative way to monitor large-scale systems for faults. The resulting filter has additional fault tolerance because it can check the health of its contituent sensors prior to deriving the top level estimate and it is easily scalable for problems which are varying in size such as collections of systems.

REFERENCES

Beard, R. V., 1971, " Failure Accommodation in Linear Systems through Self-Reorganization," Ph.D. Thesis, Massachusetts Institute of Technology.

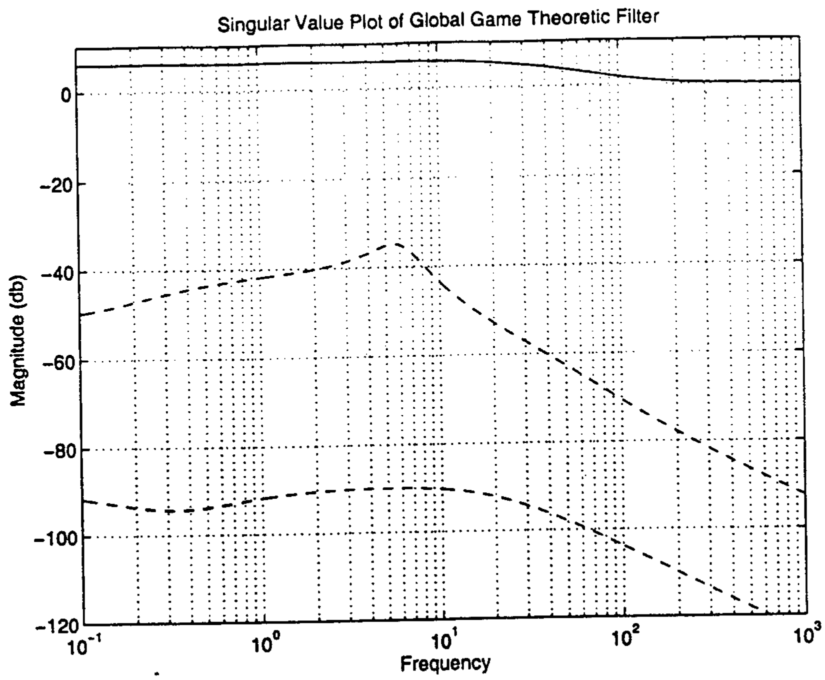

Figure 4. Platoon Example - Signal Transmission in the Global Detection Filter (position sensor fault transmission shown with solid line, nuisance fault transmission shown with dashed line)

Bowman, G. A. and Speyer, J. L., 1987, "Detection Filters for Missile Tracking," in: *Proceedings of the 1987 Conference on Guidance, Navigation, and Control*, AIAA, pp. 570–578, AIAA Paper No. 87-2383.

Chung, W. H., 1997, "Game Theoretic and Decentralized Estimation for Fault Detection," Ph.D. Thesis, University of California, Los Angeles.

Chung, W. H. and Speyer, J. L., 1995, "A General Framework for Decentralized Estimation," in: *Proceedings of the 1995 American Control Conference*, ACC, Seattle, WA.

Chung, W. H. and Speyer, J. L., 1998, "A Game Theoretic Fault Detection Filter," *IEEE Transactions on Automatic Control*, Vol. AC-43, No. 2.

Ding, X. and Frank, P. M., 1989, "Fault Detection via Optimally Robust Detection Filters," in: *Proceedings of the 28th Conference on Decision and Control*, IEEE, Tampa, FL, pp. 1767–1772.

Douglas, R. K., 1993, "Robust Detection Filter Design," Ph.D. Thesis, University of Texas at Austin.

Douglas, R. K., Speyer, J. L., Mingori, D. L., Chen, R. H., Malladi, D. P., and Chung, W. H., 1995, "Fault Detection and Identification with Application to Advanced Vehicle Control Systems," Research Report UCB-ITS-PRR-95-26, California PATH.

Douglas, R. K., Speyer, J. L., Mingori, D. L., Chen, R. H., Malladi, D. P., and Chung, W. H., 1996, "Fault Detection and Identification with Application to Advanced Vehicle Control: Final Report," Research Report UCB-ITS-PRR-96-25, California PATH.

Green, M. and Limebeer, D. J., 1995, *Linear Robust Control*, Prentice-Hall.

Jang, J. and Speyer, J. L., 1994, "Decentralized Game-Theoretic Filters," in: *Proceedings of the 1994 American Control Conference*, ACC, Baltimore, MD, pp. 3379–3384.

Kalman, R., 1964, "When is a Linear Control System Optimal?" *Transactions of the ASME, Journal of Basic Engineering, Series D*, Vol. 86, pp. 51–50.

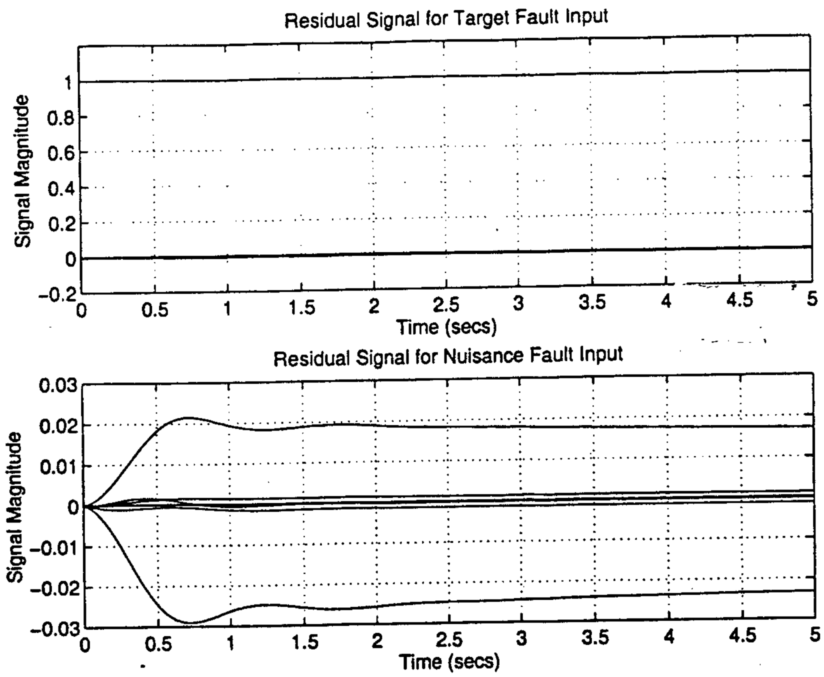

Figure 5. Platoon Example - Failure Signal Response of the Decentralized Fault Detection Filter (Nuisance Fault is a Step Failure in the Longitudinal Accelerometer on Car # 1)

Kerr, T., 1985, "Decentralized Filtering and Redundancy Management/Failure Dectection for Multisensor Integrated Naviagtion Systems," in: *Proceedings of the ION National Technical Meeting*, Institute of Navigation.

Massoumnia, M.-A., 1986, "A Geometric Approach to the Synthesis of Failure Detection Filters," *IEEE Transactions on Automatic Control*, Vol. AC-31, No. 9, pp. 839–846.

Massoumnia, M.-A., Verghese, G. C., and Willsky, A. S., 1989, "Fault Detection and Identification," *IEEE Transactions on Automatic Control*, Vol. AC-34, No. 3, pp. 316–321.

Patton, R. and Chen, J., 1992, "Robust Fault Detection of Jet Engine Sensor Systems Using Eigenstructure Assignment," *Journal of Guidance, Control, and Dynamics*, Vol. 15, No. 6, pp. 1491–1497.

Patton, R., Chen, J., and Millar, J., 1991, "A Robust Disturbance Decoupling Approach to Fault Detection in Process Systems," in: *Proceedings of the 30th Conference on Decision and Control*, IEEE, Brighton, UK, pp. 1543–1548.

Rhee, I. and Speyer, J. L., 1991, "A Game Theoretic Approach to a Finite-Time Disturbance Attenuation Problem," *IEEE Transactions of Automatic Control*, Vol. AC-36, No. 9, pp. 1021–1032.

Speyer, J. L., 1979, "Computation and Transmission Requirements for a Decentralized Linear-Quadratic-Gaussian Control Problem," *IEEE Transactions on Automatic Control*, Vol. AC-24, No. 2, pp. 266–269.

White, J. E. and Speyer, J. L., 1987, "Detection Filter Design: Spectral Theory and Algorithms," *IEEE Transactions on Automatic Control*, Vol. AC-32, No. 7, pp. 593–603.

Willsky, A. S., Bello, M. G., Castanon, D. A., Levy, B. C., and Verghese, G. C., 1982, "Combining and Updating of Local Estimates and Regional Maps Along Sets of One-Dimensional Tracks," *IEEE Transactions on Automatic Control*, Vol. AC-27, No. 4, pp. 799–813.

Wolfe, J. D., Chichka, D. F., and Speyer, J. L., 1996, "Decentralized Controllers for Unamanned Aerial Vehicle Flight," in: *Proceedings of the 1996 Conference on Guidance, Navigation and Control*, AIAA, San Diego, CA, AIAA Paper No. 96-3833.